Forecasting Cloud Cover and Atmospheric Seeing for Astronomical Observing: Application and Evaluation of the Global Forecast System

Q.-Z. YE,
Department of Atmospheric Sciences, School of Environmental Science and Engineering, Sun Yat-sen University, Guangzhou, China; tom6740@gmail.com

Received 2010 April 14; accepted 2010 November 16; published 2011 January 5

ABSTRACT. To explore the issue of performing a noninteractive numerical weather forecast with an operational global model to assist with astronomical observing, we use the Xu-Randall cloud scheme and the Trinquet-Vernin AXP seeing model with the global numerical output from the Global Forecast System (GFS) to generate 3–72 hr forecasts for cloud coverage and atmospheric seeing, and we compare them with sequence observations from nine sites from different regions of the world with different climatic backgrounds in the period from 2008 January to 2009 December. The evaluation shows that the proportion of prefect forecast of cloud cover forecast varies from ∼50% to ∼85%. The probability of cloud detection is estimated to be around ∼30% to ∼90%, while the false alarm rate is generally moderate and is much lower than the probability of detection in most cases. The seeing forecast has a moderate mean difference (absolute mean difference <0.3″ in most cases) and rms error (0.2″–0.4″ in most cases), compared with the observation. The probability of forecast with <30% error varies between 40% and 50% for the entire-atmosphere forecast and between 30% and 50% for the free-atmosphere forecast for almost all sites, which is in the better cluster among major seeing models. However, the forecast errors are quite large for a few particular sites. Further analysis suggests that the error might primarily be caused by the poor capability of the GFS/AXP model to simulate the effect of turbulence near ground and on a subkilometer scale. Overall, although the quality of the GFS model forecast may not be comparable with the human-participated forecast at this moment, our study has illustrated its suitability for a basic observing reference and has proposed its potential to gain better performance with additional efforts on model refinement.

1. INTRODUCTION

Almost all kinds of ground-based astronomical observations, especially for optical types, are extremely dependent on meteorological condition, so it is no doubt that making the meteorological prediction as precise as possible would significantly help observers to schedule the observation and improve the efficiency of telescope operation. Among all meteorological variables, cloud amount or cloud cover is apparently the dominant factor, while atmospheric/astronomical seeing, sometimes described as the Fried parameter (Fried 1965), is also important.

A climatological study to the annual values of cloud cover and/or atmospheric seeing for a proposed professional observatory is usually done in the site survey prior to the construction (see Walker 1970 and Fuentes & Muñoz-Tuñon 1990 for examples). However, to maximize the observing resources, astronomers not only need to know the approximate percentage of clear nights in a year, but they also wish to know whether the sky will be clear or not in the next few nights. In other words, they expect weather forecasts to be as accurate and precise as possible. However, since astronomical observatories are generally built in distant areas (therefore, with sparse meteorological observations available and less interest from meteorologists), special forecasts that aim at assisting astronomical observation had not been widely practiced until very recent years. Since the end of the 1990s, special forecast services and products based on mesoscale regional numeric models and/or real-time satellite images have been developed at the large professional observatories, such as Mauna Kea Weather Center (MKWC; see Businger et al. 2002 for an overview) and the nowcast model at ESO (Erasmus & Sarazin 2001).

As the operation of high-resolution numeric models would require the ability to perform speedy computation (which is usually only available at large professional observatories), attempts aimed at making direct uses of the model fields from global/continental models were carried out later, such as the Clear Sky Chart2 that has used the Canadian Meteorological Centre’s Global Environmental Multiscale (GEM)3 since 2002 (Danko 2003) and our 7Timer system4 that has used the National Centers for Environmental Prediction (NCEP) Global

---

1Current address: Room 404, 12 Huasheng Street, Guangzhou, 510620 China.

2See http://cleardarksky.com/csk/.

3See http://collaboration.cmc.ec.gc.ca/science/rpn/gef_html_public.

4See http://7timer.y234.cn.
Forecast System (GFS) since 2005 (see Sela 1980; Whitaker et al. 2008 for an overview of the model). These direct-from-model forecasts only have decent spatial and vertical resolutions (for a comparison, the spatial resolution of the GFS model is about 40 km, while the regional model operating at MKWC can reach 1 km), but it does not require heavy computation work either; the retrieval of the model fields can be done with an Internet-connected personal computer within a couple of minutes, making it the most favorable (and probably the only) choice when a speedy computer is not available and the demand on forecast precision/accuracy is not critical.

Interestingly, although these services have been put into good use by private, public, and even some professional observatories, there has been no quantitative and systematic understanding of how accurate the model fields are up to now. For example, the only reported estimation of the accuracy of cloud field forecast for a global model was done by Erasmus & Sarazin (2001) from 1992–1993, which suggested that only 15–25% of cloudy nights could be identified with the European Centre for Medium-Range Weather Forecasting model. This is age-old, considering there had been a number of significant upgrades of the global models in the following decade. The forecast for atmospheric seeing, on the other hand, is more complex, since it is related to vertical fine structure of the atmospheric column and is not directly provided as part of the output in any global model. In general, there are two tracks for atmospheric seeing forecast: nowcast track using near real-time meteorological observation profile combined with a statistical model (such as Murtagh et al. 1995) or the model track using either the derivations of Tatarski’s formula (Tatarski 1961; Coulman et al. 1988) or the numeric model proposed by Coulman et al. (1986). The first track is relatively intuitive and is accurate enough on many occasions, but it has a very short forecast range (usually less than 24 hr) and heavily depends on the availability and quality of the observational data; for the second track, one would need to divide the atmospheric column into a good number of layers to gain a numeric simulation that is close enough to the actual situation, which will again require assistance from a speedy computer. In order to solve these shortcomings, Trinquet & Vernin (2006) took advantages from both tracks and introduced a new seeing model called the AXP model. With that model, one only needs to divide the atmospheric column by a number close to that available in most global models, and the consistency from the simulation of the AXP model to the actual situation is satisfying according to the authors. Overall, these direct-from-model forecasts can be a practical solution for the observers without the ability to operate a high-precision regional model of their own, and the job to do is to assess the accuracy of these forecasts.

We organize this article as follows. In § 2, we briefly outline the technical details of the GFS model used in this study, the Xu-Randall cloud scheme used for cloud simulation, and the AXP model used to derive forecast of atmospheric seeing. In § 3, we describe the observations we used to evaluate the GFS model. Section 4 presents the details and discussions of the evaluation methodology and result, and § 5 gives the concluding remarks of this study.

2. FORECAST

2.1. The NCEP GFS Model

The GFS model provides output in two grids with different spatial resolutions: grid 003 at 1° × 1° and grid 004 at 0.5° × 0.5°. To provide the best-possible forecast, we use the latter in our study. Model outputs from the period from 2008 January 1 to 2009 December 31 at three hourly intervals for 0 < \( \tau \leq 72 \) hr at 00Z initialization are retrieved from the National Operational Model Archive & Distribution System (see Rutledge et al. 2006) for evaluation.

The GFS data set contains approximately 140 fields, supplying forecast fields for general meteorological interests (such as temperature, humidity, wind direction and speed, etc.) and for special purpose, including cloud cover fraction on different layers (low, mid, high, convective, and total atmospheric column). Although the atmospheric seeing is not among the output fields, it can be derived indirectly, as all of the required meteorological variables are given.

2.2. Cloud Scheme

In the GFS model, the cloud cover fraction for each grid box is computed using the cloud scheme presented by Xu & Randall (1996), which is shown as equation (1). In this equation, RH is the relative humidity, \( q^* \) and \( q_c \) are the saturation specific humidity, and \( q_{c\min} \) is a prescribed minimum threshold value of \( q_c \). Depending on the environmental temperature, \( q^* \) and \( q_c \) are calculated with respect to water phase or ice phase (F. Yang 2010, private communication). Cloud cover fraction can therefore be calculated for any layer as long as the RH, \( q^* \), and \( q_c \) are known and \( q_{c\min} \) is suitably prescribed. We note that the calculation is done as part of the model simulation at NCEP, so the cloud fields are used as is from the GFS data sets:

\[
C = \max \left[ \frac{2000(0.6 - q_{c\min})}{[1 - \min(\max(0, RH)|q^*/q_c^{2000(0.6 - q_{c\min})})], 0, 0} \right].
\]

The GFS model divides the whole atmospheric column into 26 layers. The total cloud cover for the entire atmospheric column is derived under the assumption that clouds in all layers are maximally randomly overlapped (Yang et al. 2005).

\[\text{See http://www.emc.ncep.noaa.gov/GFS/}.\]
\[\text{See http://www.ecmwf.int/}.\]
2.3. Seeing Model

The way atmospheric optical turbulence affects astronomical observing is theoretically described by the Kolmogorov-Tatarski turbulence model (Tatarski 1961; Roddier 1981; Tokovinin 2002), which suggested that only one parameter is needed to describe the quality of atmospheric seeing where \( \lambda \) is the wavelength associated with seeing \( ( \lambda = 5 \times 10^{-7} \text{ m in most cases}) \) and \( r_0 \) is the Fried parameter:

\[
e_0 = 0.98 \frac{\lambda}{r_0}.
\]  
(2)

The Fried parameter, \( r_0 \), is defined as follows in the direction of zenith (Coulman 1985), where \( Z_0 \) is the geopotential height of the observing site, \( C_N^2 \) is the refractive index structure coefficient indicating the strength of turbulence associated with the temperature structure coefficient \( C_T^2 \), \( P \) is the pressure in hec-topascals, and \( T \) is the temperature in degrees Kelvin:

\[
r_0 = \left[ \frac{0.423}{\lambda} \left( \frac{2\pi}{\lambda} \right)^2 \int_{Z_0}^{\infty} C_N^2 dZ \right]^{-\frac{1}{4}},
\]  
(3)

\[
C_N^2 = C_T^2 \left( \frac{7.9 \times 10^{-5} P}{T^2} \right)^2.
\]  
(4)

Therefore, with the preceding formulas and suitable inputs, we can derive the total effect of atmospheric turbulence from the integral of \( C_N^2(Z) \) for all atmospheric layers:

\[
C_T^2 = \frac{T(x) - T(x + r)}{|r|^2}.
\]  
(5)

There are several ways to derive \( C_T^2 \). A theoretical approach is shown as equation (5), where \( x \) and \( r \) are the position and separation vectors, respectively. However, in theory, \( |r| \) should be around a few tenths of a meter to precisely describe the effect of turbulence (which is not practical for model simulation, as it would require one to use about 100,000 layers for model simulation). The AXP model uses an alternative approach by considering a simple expression of \( C_T^2(h) \) as follows, where the power \( p(h) \) adjusts the amplitude of peaks and \( A(h) \) connects the level:

\[
C_T^2 = \left\langle C_T^2 \right\rangle (h) \left[ A(h) \frac{d\theta}{dz} \right]^{p(h)}.
\]  
(6)

The potential temperature \( \theta \) in equation (6) can be calculated by Poisson’s equation, where \( T \) is the absolute temperature of a parcel in degrees Kelvin and \( P \) is the pressure of that air parcel in hec-topascals:

\[
\theta = T \left( \frac{1000}{P} \right)^{0.286}.
\]  
(7)

The authors of the AXP model (Trinquet & Vernin (2006)) then statistically determined the values of the three coefficients, \( \left\langle C_T^2 \right\rangle(h), A(h), \) and \( p(h) \), by a vertical spacing of 1 km up to an altitude of 30 km, based on airborne observations from 162 flights at nine sites during 1990–2002. On the other hand, the coordinate system adopted by the GFS model is the \( P \)-coordinate system, which divides the atmospheric column by pressure. The vertical spacing of the GFS model over low- and midlevel atmospheres is around 1 km and is roughly compatible with the \( Z \)-coordinate system adopted by the AXP model; however, the former becomes too sparse at high-level atmosphere (as illustrated in Table 1). To solve this problem, we set up a degeneracy scheme to allow using the AXP model coefficients in the \( P \)-coordinate system by weighting the values according to the correlation between atmospheric pressure and altitude defined by the U.S. Standard Atmosphere, 1976.\(^7\) The transformed coefficients for each pressure layer are given in Table 2. As the output of the GFS model provides fields of temperature and atmospheric pressure for each of its vertical layer, we can align

\(^7\) See http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19770009539_1977009539.pdf.
the fields with equations (2), (3), (4), (6), and (7) to derive final $\epsilon_0$ in the form of arcseconds. We refer to this hybrid model as the GFS/AXP model hereafter. Trinquet & Vernin (2006) reported an accuracy of 58% of the original AXP model forecasts, with error within $\pm 30\%$ of observations. However, as we have modified the model layers to fit the GFS model, some additional errors may have been induced. One may expect two possible sources of errors caused by the degeneracy. The first kind of possible error comes from the layers higher than 10 km. As the GFS model layers with altitude of $>10$ km have a vertical thickness significantly larger than 1 km, one may suggest that some performance may be lost due to data roughness. However, we argue that the loss of performance for this reason should be minimal, as the high-level atmosphere is much less active than the mid- or low-level atmospheres and contributes minimal turbulence to seeing, compared

| TABLE 2 | COEFFICIENTS FOR AXP MODEL AND CORRESPONDING GFS MODEL LAYER AND WEIGHT |
|---------|--------------------------|-----------------|-----------------|-----------------|-----------------|
| AXP layer | GFS layer and weight | $P$ (hPa) | $p(h)$ | $A(h)$ | $(C_{ij}^2)/(h)$ |
| 0–50 m above ground | 2 m above ground to 0.995σ | $P(h) = 3$ hPa | 0.5 | 1.6E+2 | 1.4E–2 |
| 50–100 m above ground | 2 m above ground to 0.995σ | $P(h) = 9$ hPa | 0.5 | 1.6E+2 | 4.8E – 4(h/100)$^{–0.6}$ |
| 100–1000 m above ground | 0.995σ to 30 hPa above ground | $P(h) = 60$ hPa | 0.5 | 1.6E+2 | 3.4E – 5(h/1000)$^{1.1}$ |
| 1–2 km | 900–850 hPa (50%) | 845 hPa | 0.3 | 1.8E+3 | 5.2E–5 |
| 1–2 km | 850–800 hPa (40%) | 845 hPa | 0.3 | 1.8E+3 | 5.2E–5 |
| 1–2 km | 800–750 hPa (10%) | 845 hPa | 0.3 | 1.8E+3 | 5.2E–5 |
| 2–3 km | 750–700 hPa (50%) | 745 hPa | 1.3 | 5.6E+2 | 4.2E–5 |
| 3–4 km | 700–650 hPa (60%) | 655 hPa | 1.7 | 4.6E+2 | 2.6E–5 |
| 3–4 km | 650–600 hPa (40%) | 655 hPa | 1.7 | 4.6E+2 | 2.6E–5 |
| 4–5 km | 650–600 hPa (40%) | 575 hPa | 1.7 | 3.8E+2 | 2.5E–5 |
| 4–5 km | 600–550 hPa (70%) | 575 hPa | 1.7 | 3.8E+2 | 2.5E–5 |
| 5–6 km | 550–500 hPa (60%) | 505 hPa | 2.6 | 2.6E+2 | 1.8E–5 |
| 5–6 km | 500–450 hPa (40%) | 505 hPa | 2.6 | 2.6E+2 | 1.8E–5 |
| 6–7 km | 500–450 hPa (30%) | 440 hPa | 1.1 | 4.6E+2 | 1.5E–5 |
| 6–7 km | 450–400 hPa (70%) | 440 hPa | 1.1 | 4.6E+2 | 1.5E–5 |
| 7–8 km | 450–400 hPa (20%) | 380 hPa | 0.8 | 6.8E+2 | 1.6E–5 |
| 7–8 km | 400–350 hPa (80%) | 380 hPa | 0.8 | 6.8E+2 | 1.6E–5 |
| 8–9 km | 400–350 hPa (10%) | 330 hPa | 0.6 | 1.0E+3 | 1.6E–5 |
| 8–9 km | 350–300 hPa (90%) | 330 hPa | 0.6 | 1.0E+3 | 1.6E–5 |
| 9–10 km | 350–300 hPa (20%) | 285 hPa | 0.3 | 2.2E+3 | 1.9E–5 |
| 9–10 km | 300–250 hPa (80%) | 285 hPa | 0.3 | 2.2E+3 | 1.9E–5 |
| 10–11 km | 300–250 hPa (40%) | 245 hPa | 0.5 | 6.8E+2 | 2.6E–5 |
| 10–11 km | 250–200 hPa (60%) | 245 hPa | 0.5 | 6.8E+2 | 2.6E–5 |
| 11–12 km | 250–200 hPa (80%) | 210 hPa | 0.7 | 3.2E+2 | 3.4E–5 |
| 11–12 km | 200–150 hPa (20%) | 210 hPa | 0.7 | 3.2E+2 | 3.4E–5 |
| 12–13 km | 200–150 hPa (100%) | 177 hPa | 0.6 | 2.2E+2 | 4.4E–5 |
| 13–14 km | 200–150 hPa (50%) | 150 hPa | 0.1 | 8.3E+3 | 4.8E–5 |
| 13–14 km | 150–100 hPa (50%) | 150 hPa | 0.1 | 8.3E+3 | 4.8E–5 |
| 14–15 km | 150–100 hPa (100%) | 126 hPa | 0.2 | 1.0E+3 | 5.5E–5 |
| 15–16 km | 150–100 hPa (80%) | 105 hPa | –0.4 | 3.2E+1 | 6.5E–5 |
| 15–16 km | 100–70 hPa (20%) | 105 hPa | –0.4 | 3.2E+1 | 6.5E–5 |
| 16–17 km | 100–70 hPa (100%) | 88 hPa | –0.3 | 1.5E+1 | 8.3E–5 |
| 17–18 km | 100–70 hPa (70%) | 72 hPa | 2.5 | 5.6E+1 | 1.1E–4 |
| 17–18 km | 70–50 hPa (30%) | 72 hPa | 2.5 | 5.6E+1 | 1.1E–4 |
| 18–19 km | 70–50 hPa (100%) | 59 hPa | –0.9 | 2.6E+1 | 1.1E–4 |
| 19–20 km | 70–50 hPa (30%) | 48 hPa | 3.3 | 3.8E+1 | 9.5E–5 |
| 19–20 km | 50–30 hPa (70%) | 48 hPa | 3.3 | 3.8E+1 | 9.5E–5 |
| 20–21 km | 50–30 hPa (100%) | 39 hPa | –1.0 | 2.6E+1 | 8.2E–5 |
| 21–22 km | 50–30 hPa (60%) | 31 hPa | 1.5 | 4.6E+1 | 7.4E–5 |
| 21–22 km | 30–20 hPa (40%) | 31 hPa | 1.5 | 4.6E+1 | 7.4E–5 |
| 22–23 km | 30–20 hPa (100%) | 24 hPa | –1.9 | 3.2E+1 | 7.5E–5 |
| 23–24 km | 30–20 hPa (30%) | 19 hPa | 1.3 | 4.6E+1 | 8.7E–5 |
| 23–24 km | 20–10 hPa (70%) | 19 hPa | 1.3 | 4.6E+1 | 8.7E–5 |
| 24–25 km | 20–10 hPa (100%) | 15 hPa | 1.1 | 3.8E+1 | 1.1E–4 |
| 25–26 km | 20–10 hPa (100%) | 11 hPa | 1.5 | 3.8E+1 | 1.3E–4 |
so a vertical spacing degeneration from 1 km to 2 km at high-level atmosphere is unlikely to induce error of significance. On the other hand, the degeneracy in planetary boundary layer (PBL), which serves as the second possible source of error, may induce something large. PBL is a layer that is directly influenced by the atmosphere-ground interaction, and it has been shown that the PBL turbulence contributes a major part to the seeing during the night (Abahamid et al. 2004). To improve the model performance at PBL, the AXP model divides the PBL into three sublayers: 0–50 m above ground, 50–100 m above ground, and 100–1000 m above ground. However, the GFS model only provides three near-ground fields, which are at 2 m above ground, 0.995σ (~100 m above ground), and 30 hPa above ground (~400 m). We have to use identical $d\theta/dz$ computed from 2 m above ground, 0.995σ for layers of 0–50 m and 50–100 m, and the $d\theta/dz$ computed from 0.995σ and 30 mb above ground for layers of 100–1000 m. The ignorance of data at 50 m would almost certainly induce some error. To work around this issue and allow a direct assessment of the GFS/AXP model, we also generate a seeing forecast for free atmosphere only—i.e., with PBL (<100 m in our study) excluded—for our evaluation. The free-atmosphere forecast will be compared with the observations from a multi-aperture scintillation sensor (MASS), which is designed to measure the seeing in free atmosphere (Tokovinin 2002).

3. OBSERVATION

We collect cloud cover and seeing observations for a total of nine sites from central/east Asia, Hawaii, and Central/South America in the period from 2008 January to 2009 December.

### Table 3

| Site            | Location | Elevation (m) | GFS grid elevation | Climate | Data type | Availability | PBL top | $P(\theta)$ |
|-----------------|----------|---------------|--------------------|---------|-----------|--------------|---------|-------------|
| Paranal         | −70.40, −24.63 | 2635 m        | 919 m              | Arid    | Cloud, DIMM | 2008 Jan 1–2009 Dec 31 | 3700 m | 734 hPa     |
| Mauna Kea       | −155.48, +19.83 | 4050 m        | 896 m              | Highlands | DIMM, MASS | 2008 Jan 1–2008 May 31 | 5100 m | 612 hPa     |
| San Pedro Mártr | −115.47, +31.05 | 2830 m        | 1038 m             | Arid    | DIMM, MASS | 2008 Jan 1–2008 Aug 31 | 3800 m | 716 hPa     |
| Cerro Tololo    | −70.80, −30.17 | 2200 m        | 926 m              | Arid    | MASS       | 2008 Jan 1–2009 Dec 31 | 3200 m | 775 hPa     |
| Nanshan         | +87.18, +43.47 | 2080 m        | 2233 m             | Semiarid | Cloud      | 2008 Jan 1–2009 Jul 4  | ...    | ...         |
| Lulin           | +120.87, +23.47 | 2862 m        | 1345 m             | Humid   | subtropical | 2008 Jan 1–2009 Jul 31 | ...    | ...         |
| Cerro Armazones | −70.20, −24.60 | 3064 m        | 1997 m             | Arid    | DIMM, MASS | 2008 Jan 1–2009 Sep 31 | 4100 m | 695 hPa     |
| Cerro Pachón    | −70.73, −30.23 | 2715 m        | 2706 m             | Arid    | DIMM, MASS | 2008 Jan 1–2009 Sep 31 | 3700 m | 727 hPa     |
| Cerro Tolonchar | −67.98, −23.93 | 4480 m        | 3626 m             | Highlands | DIMM, MASS | 2008 Jan 1–2008 Sep 30 | 5500 m | 579 hPa     |

1 PBL top stands for the upper limit of the PBL adopted in the GFS/AXP model.
2 $P(\theta)$ is the atmospheric pressure at the height of the site used in GFS/AXP model, calculated under standard atmosphere.
3 Ground-based seeing observation from 2008 October 31 to 2009 April 29 only.

8 Study by Li et al. (2003) suggests the atmosphere beyond 100 hPa (~15,000 m) contributes less than 0.01" to the seeing.
9 We do not exclude the 100 m–1000 m region in our free-atmosphere seeing forecast, because the MASS instrument used for measurement of free atmosphere seeing considers an altitude of 500 m above ground as the PBL top limit and includes measurements of layers at an altitude as low as 500 m and 1000 m above ground.

### Table 4

| Category | Paranal | Nanshan | Lulin |
|----------|---------|---------|-------|
|          | 30% Threshold |         |       |
| H        | 73      | 715     | 682   |
| F        | 443     | 482     | 860   |
| M        | 71      | 38      | 14    |
| Z        | 1411    | 241     | 94    |
| Total n  | 1998    | 1476    | 1650  |
|          | 50% Threshold |         |       |
| H        | 62      | 684     | 670   |
| F        | 353     | 386     | 813   |
| M        | 82      | 69      | 26    |
| Z        | 1501    | 337     | 141   |
| Total n  | 1998    | 1476    | 1650  |
|          | 80% Threshold |         |       |
| H        | 44      | 576     | 642   |
| F        | 225     | 232     | 734   |
| M        | 100     | 177     | 54    |
| Z        | 1629    | 491     | 220   |
| Total n  | 1998    | 1476    | 1650  |
The respective information of each site, including the PBL top limit and \( P(h) \), used in the GFS/AXP model is listed in Table 3.

As these observations are all made in sequence with a sampling frequency of around 1–1.5 min, except Nanshan and Lulin (which will be dealt with separately and will be described subsequently), they are first processed to match the time interval of the GFS model output (which is 3 hr). To ensure that the observation is representative in the corresponding interval, we set a minimum data points of 100 for each interval. A sampling frequency of 1–1.5 minutes corresponds 120–180 data points per 3 hr, so the minimum threshold of 100 is reasonable.

As all seeing observations are obtained either by a differential image motion monitor (DIMM) or MASS, they can be used without further reduction, since they give the measurement of \( \epsilon_0 \) directly. On the other hand, the cloud cover observations from Paranal, Nanshan, and Lulin are obtained with different tools, so they must be reduced to the same definition with the GFS model output before comparing them with the latter. The reduction procedure is described later.

**Paranal.**—The cloud sensor installed at Paranal determines the sky condition by measuring the flux variation of a star. The sensor graph will suggests a possible cloudy condition when the rms of the flux variation is larger than 0.02.\(^{10}\)

**Nanshan.**—For Nanshan, the operation log is used to verify the sky condition. The operation log includes the image sequence log and observer’s notes. We divide each night into two parts: evening and morning. A part with roughly >50% observable time (with images taken and indication of good observing condition from the observer) would be marked as clear; otherwise, it would be marked as cloudy.

**Lulin.**—A diffraction-limited Boltwood loud Sensor is installed at Lulin to produce sequence observations. The cloud sensor determines the sky condition by comparing the temperature of the sky with ambient ground-level temperature, and a difference threshold of \(-25^\circ C\) is set to distinguish clear and cloudy conditions.\(^{11}\)

### 4. Evaluation

#### 4.1. Evaluation of Cloud Cover Forecast

The preliminary result from our companion study suggests that roughly 70% of cloud cover forecasts from the GFS model can achieve an error of less than 30% (Ye & Chen 2010) at 0 hr < \( \tau \) ≤ 72 hr at the grid with spacing of ~300 km. However, the cloud cover observations used in this study are all made at *single geographic points* and are *categorical* (either clear or cloudy), rather than over a grid area and being quantitative. So in order to make comparison between the forecasts and the observations, we first need to simplify the forecast into two categories by setting a spatial threshold that divides cloudy and clear situations. To examine the result sensitivity with different thresholds, we set the threshold at 30%, 50%, and 80%, respectively. Considering that most astronomical observatories are placed at areas with higher chances of being clear rather than cloudy, we should focus on the cloudy event in our study, rather than on the clear event. Thus, the cloudy event is set to be the “yes” statement, and clear is set to be the “no” statement.

We use four indicators to evaluate the performance of the forecast:\(^{12}\) proportion of perfect forecasts (PPF), probability of detection (POD), false alarm rate (FAR), and frequency bias index (FBI). Let \( H \) denote hits (forecasted and observed), let \( F \) denote false alarms (forecasted but not observed), let \( M \) denote missed (not forecasted but observed), and let \( Z \) denote non-forecasted and not-observed situations; we then have

\[
\text{PPF} = \frac{H + Z}{H + F + M + Z},
\]

\[
\text{POD} = \frac{H}{H + M},
\]

\[
\text{FAR} = \frac{F}{F + Z},
\]

and

\[
\text{FBI} = \frac{H + F}{H + M}.
\]

---

\(^{11}\) See http://www.cyanogen.com/downloads/files/claritymanual.pdf.

\(^{12}\) A more detailed technical document about the application of these four indicators in meteorology may be found at http://www.ecmwf.int/products/forecasts/guide/Hit_rate_and_False_alarm_rate.html.
The $H$, $F$, $M$, and $Z$ numbers for each of the three sites are listed in Table 4, and the comparison results between the forecast and observation are shown in Table 5.

As the result from our companion study has implied that the accuracy declination with the increase of $\tau$ is small (rms error [rmse] varies within 5% for $\tau$ up to 72 hr) for the GFS model, we see no need to list percentages for each night separately. The percentages given in Table 5 are the average of $0 \text{ hr} < \tau \leq 72 \text{ hr}$, where the uncertainty ranges are denoted by the nights with highest/lowest percentages. The large uncertainty of the FBI of Paranal is caused by small dominators, since the number of false alarms is 5–10 times higher than the number of misses.

### Table 6: Mean Difference and RMSE of Seeing Forecast for Sample Sites

| Site                  | Entire atmosphere mean difference $^a$ | RMSE  | Sample $n$ | Free atmosphere mean difference $^a$ | RMSE  | Sample $n$ |
|-----------------------|----------------------------------------|-------|------------|--------------------------------------|-------|------------|
| Paranal               | $-0.09^\prime$                        | 0.36\arcsec | 3630       | $...$                                | $...$ | $...$      |
| Mauna Kea             | $-0.15^\prime$                        | 0.26\arcsec | 34         | $-0.32^\prime$                      | 0.42\arcsec | 42         |
| San Pedro Mártir      | $+0.01^\prime$                        | 0.45\arcsec | 326        | $-0.10^\prime$                      | 0.22\arcsec | 118        |
| Cerro Tololo          | $...$                                  | $...$      | $...$      | $-0.24^\prime$                      | 0.34\arcsec | 640        |
| Cerro Armazones       | $+0.20^\prime$                        | 0.34\arcsec | 776        | $-0.18^\prime$                      | 0.27\arcsec | 878        |
| Cerro Pachón          | $+0.46^\prime$                        | 0.50\arcsec | 738        | $-0.14^\prime$                      | 0.27\arcsec | 928        |
| Cerro Tolonchar       | $+0.29^\prime$                        | 0.40\arcsec | 298        | $-0.26^\prime$                      | 0.33\arcsec | 92         |

$^a$Calculated by the mean of forecast minus the mean of observation.

![Fig. 1](image1.png) Absolute forecast mean errors and the height differences between actual height and GFS grid for all sites for entire atmosphere (top) and free atmosphere (bottom). Abbreviations are P—Paranal, SPM—San Pedro Mártir, MK—Mauna Kea, CO—Cerro Tololo, CA—Cerro Armazones, CT—Cerro Tolonchar, and CP—Cerro Pachón.

![Fig. 2](image2.png) Absolute forecast mean error and the altitude of the site for all sites with DIMM observation (top) and MASS observation (bottom). Abbreviations are P—Paranal, SPM—San Pedro Mártir, MK—Mauna Kea, CO—Cerro Tololo, CA—Cerro Armazones, CT—Cerro Tolonchar, and CP—Cerro Pachón.
Although PPF is a biased score, since it is strongly influenced by the more common category, it gives a rough indication about the typical percentage of correct forecasts for one site. A clear feature revealed by Table 5 is the climate dependence of the PPF: arid sites tend to have higher PPF, while humid sites tend to have lower. Generally speaking, the PPF varies from 50% to 85% for the three sites in our sample. As the sites have covered both the extremes of climate type (from arid and humid), it can be considered that this result is relatively representative.

The result revealed by POD is also encouraging. Even for Paranal, the site with an annual cloudy percentage as low as 10% (Ardeberg et al. 1990), the POD can still vary between 20 and 60%, which is better than the 15–25% percentage reported by (Erasmus & Sarazin 2001). We also note that the forecast FAR for Paranal and Nanshan is less than half of the forecast POD, which is far below the human-participated forecasts (for reference, the MKWC forecaster, which can reach a FAR as low as 1%)\(^{11}\), but is still reasonable for a basic observing reference. We note that the forecast FAR for Lulin is very large; however, considering that Lulin is surrounded by a very complex terrain, where the elevation variation in its assigned grid box is as large as 3500 m, it is not unexpected for a nonideal FAR. Finally, the FBIs for all sites are larger than 1, suggesting the GFS model tends to make more false alarms rather than misses.

4.2. Evaluation of Seeing Forecast

The evaluation of the seeing forecast is relatively easier than that of cloud cover forecast, since the forecasts and observations are in the same definition, and we can simply compare the mean and rmse of the difference between the forecasts and observations (Table 6). The mean errors are computed by averaging the differences between instantaneous forecasts and observations. As the increase of error is small ($\Delta$rmse < $\sim$0.05") for $\tau$ up to 72 hr, we combine forecasts and observations from all three nights into a whole for our study.

As shown in Table 6, either the entire-atmosphere forecast or free-atmosphere forecast has moderate error, compared with the observation; the former tends to slightly overestimate the value and the latter tends to underestimate the value. However, the tendency is severe for the cases such as Cerro Pachón, for which the mean difference reads $+0.46"$. A possible explanation is that many observatories are located in mountainous areas and are actually much higher than the height of their assigned grids in GFS models (the height differences vary from 0–4000 m among our sample, as shown in Table 3); this may produce error in PBL seeing estimation. However, we find no substantial support for this assumption, as there is no significant tendency between the forecast bias and the height difference (Fig. 1). Dramatically, the site with the smallest height difference (Cerro Pachón, with a height difference of 9 m) has the largest mean error ($+0.46"$).

---

\(^{11}\) See http://mkwc.ifa.hawaii.edu/forecast/mko/stats/index.cgi?night = 1&fcster = fcsts&var = fog&cut = 2.
FIG. 5.—Forecast-by-observation figure of entire-atmosphere seeing forecast for Paranal (P), Mauna Kea (MK), San Pedro Mártir (SPM), Cerro Armazones (CA), Cerro Pachón (CP), and Cerro Tolonchar (CT). Dashed line corresponds to the ideal case (slope = 1), and dotted lines are 30% error uncertainty from the ideal case.
Fig. 6.—Forecast-by-observation figure of free-atmosphere seeing forecast for Mauna Kea (MK), San Pedro Mártir (SPM), Cerro Tololo (CO), Cerro Armazones (CA), Cerro Pachón (CP), and Cerro Tolonchar (CT). Dashed line corresponds to the ideal case (slope = 1), and dotted lines are 30% error uncertainty from the ideal case.
Another equally plausible explanation is the combined effect of the poor consistency between the GFS/AXP model versus the actual $C_n^2$ variance in PBL and error induced by the layer degeneration of the AXP model for high-altitude areas. In fact, we do find a weak-bias tendency for the free-atmosphere forecast, as shown in Figure 2; the forecast for low-altitude sites tends to be better, and all sites below 3000 m have an absolute mean forecast error below 0.25". This feature fits the fact that the airborne data used to determine the coefficients in the AXP model only include sites with altitudes up to 2835 m. But generally speaking, no substantial correlation between bias tendency and a unique geographic modeling factor can be identified; therefore, we can only suggest that the bias is caused by a combination effect of some factors.

To give a more comprehensive understanding of the quality of the GFS/AXP forecast, we plot the cumulative distribution of the relative forecast error for each site in Figure 3 (forecast for the entire atmosphere) and Figure 4 (forecast for the free atmosphere only). Let $\protect\epsilon_0$ be the forecasted seeing value and let $\epsilon_o$ be the observed seeing value; the relative forecast error $E(\epsilon_0)$ is then calculated by

$$E(\epsilon_0) = \left| \frac{\epsilon_0 - \epsilon_o}{\epsilon_o} \right|.$$  

From Figures 3 and 4, we can identify that the probabilities of producing forecast with $<30\%$ error concentrate in $40$–$50\%$ for the entire-atmosphere seeing forecast and $30$–$50\%$ for the free-atmosphere seeing forecast. By contrast, Trinquet & Vernin (2006) gave a probability of $58\%$ and $50\%$ for the original AXP model to produce an entire-atmosphere or free-atmosphere forecast with the same quality. Generally speaking, our result is in rough agreement with theirs, although the model’s performance is rather unsatisfying on a few particular sites.

In addition to the summarizing table and cumulative distribution figures, we also present the forecast-observation distribution for each site in Figure 5 and Figure 6. We can see that although the statistical values might be satisfying, the forecast-observation distribution figures imply that the correlation between forecast and observation is poor. This is not unexpected, since the study of Cherubini et al. (2009) at MKWC has suggested that one may achieve a good approximation of the actual condition with $\sim$80 vertical layers up to 10 hPa, and so the $\sim$15 layers used in our model may be far from sufficient for a well-correlated distribution. Our assumption is reinforced by the fact that forecasts for free-atmosphere seeing are generally concentrated in a narrow region, despite their mean being close to that of the observation, as shown in Figure 6. This phenomenon implies that turbulence over the subkilometer scale in the vertical direction might be the major contributor to a bad seeing condition in the free-atmosphere region.

Finally, we compare our result with several major models/forecasters (Table 7). The rmse and $<30\%$ error probabilities for the original AXP model, Vernin-Tatarski model, $C_n^2$/seeing median, and seeing mean are derived by Trinquet & Vernin (2006) with their experimental profile observations. The MKWC Weather Research and Forecasting (WRF) model is initiated with the GFS model output, but the simulation eventually derives output with a final grid spacing of 1 km and a vertical layer number of 40, resulting in a forecast rmse at 0.36" for the first night. As revealed by Table 7, the rmse uncertainty ranges of GFS/AXP model fall between the mean of MKWC forecaster and the original AXP model, while the $<30\%$ error probability is in the better cluster in all models under most cases. Interestingly, a direct comparison of the GFS/AXP model with the MKWC forecaster over the seeing forecast for the same site (Mauna Kea) even slightly favors the former (0.26" versus 0.28" for three-night mean). In short, this result has highlighted the potential of the GFS/AXP model to be a competitive forecast tool once the layer degeneracy and high-altitude issues are solved.
5. CONCLUSION

We have carried out a comprehensive study over the topic of performing automatic numeric forecast of cloud cover and atmospheric seeing with the Global Forecast System, an operational global model. Sequence observations on cloud cover and atmospheric seeing from nine sites from different regions of the world with different climatic backgrounds in the period from 2008 January to 2009 December are used to evaluate the forecast. Although the performance of the model forecast may not be comparable with the human-participated forecast, our study has shown the forecast to be acceptable for a basic observing reference. Our study has also revealed the possibility of gaining better performance from the model with additional efforts on model refinement.

For the cloud cover forecast, we have found that the proportion of perfect forecasts varies from ~50% to ~85% for all three sites we evaluated, including a site located in a subtropical region with a very humid climate (Lulin). In particular, we have found that the model is capable of detecting a significant amount of occurring clouds, and the false alarm rate is moderate. The probability of detection is still measured to be 20–60%, even for sites with very low cloud probability (Paranal).

For the atmospheric seeing forecast, we adopted the AXP model introduced by Trinquet & Vernin (2006). We found that the forecast for the entire atmosphere tends to slightly overestimate the seeing, while the free-atmosphere forecast tends to underestimate it. The rmse for free-atmosphere seeing is smaller (0.22–0.42") than that of the entire atmosphere (0.26–0.50"), but both values can indicate a decent quality of the forecast, compared with the other major models. Further analysis suggests that a major contributor to the forecast error might be the layer degeneracy issue of the hybrid GFS/AXP model. On the other hand, the probability of GFS/AXP forecasts with <30% error varies between 40–50% for the entire-atmosphere forecast and 30–50% for free-atmosphere forecast in most cases, which is in the better cluster among major seeing models. To conclude, our study has suggested a decent performance of the GFS/AXP model that is suitable for basic observing reference. Our study has also suggested that the model has the potential to become a rather useful forecast tool with additional efforts on model refinement.

The author wishes to thank all the people who have given help to this work, particularly to an anonymous reviewer for very constructive comments that led to an advance of this work and to Sisi Chen, Shaojia Fan, Linjiong Zhou, and Johnson Lau for their help from various aspects. The author is very grateful to Chenzhou Cui on behalf of the Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST) team at National Astronomical Observatory, Chinese Academy of Sciences, who made this work possible by providing excellent long-term resources of server and computing for our 7Timer system since 2005. The author also wishes to thank MountナンサンXingming Observatory, National Central University Lulin Observatory, Cerro Tololo Inter-American Observatory, European Southern Observatory, and the Thirty Meter Telescope (TMT) Site Testing Project for providing and/or granting permission to access and use the astrometeorological observation data from Mountナンサン, Lulin, Cerro Pachón, Paranal, and the TMT test sites. In addition, the author would like to thank Xing Gao (No. 1 Senior High School of Ürümqi/Xingming Observatory), Hung-Chin Lin (National Central University/Lulin Observatory), Andrei Tokovinin (Cerro Tololo Inter-American Observatory), Edison Bustos (Cerro Tololo Inter-American Observatory), and Felipe Daruich (Gemini Observatory) for providing assistance with obtaining the observational data.

REFERENCES

Abahamid, A., et al. 2004, A&A, 422, 1123
Ardeberg, A., Lundstrom, I., & Lindgren, H. 1990, A&A, 230, 518
Businger, S., McLaren, R., Ogasawara, R., Simons, D., & Wainscoat, R. J. 2002, Bull. Am. Meteorol. Soc., 83, 858
Cherubini, T., Businger, S., & Lyman, R. 2009, in Optical Turbulence: Astronomy Meets Meteorology, Proc. Optical Turbulence Characterization for Astronomical Applications (Hackensack: World Scientific), 196
Coulman, C. E. 1985, ARA&A, 23, 19
Coulman, C. E., André, J.-C., Lacarrère, P., & Gillingham, P. R. 1986, PASP, 98, 376
Coulman, C. E., Vernin, J., Coqueugniot, Y., & Caccia, J.-L. 1988, Appl. Opt., 27, 155
Danko, A. 2003, S&T, 105, 62
Erasmus, D. A., & Sarazin, M. 2001, Proc. SPIE, 4168, 317
Fried, D. 1965, J. Opt. Soc. Am., 55, 1427
Fuentes, F. J., & Muñoz-Tuñón, C. 1990, Ap&SS, 171, 267
Li, S.-X., Fu, Y.-F., Huang, Y.-L., Li, J.-G., & Mao, J.-T. 2003, Acta Astron. Sinica, 44, 431

Ye, Q.-Z., & Chen, S.-S. 2010, J. Atmos. Sci., submitted

Murtagh, F., Assem, A., & Sarazin, M. 1995, PASP, 107, 702
Roddier, F. 1981, Prog. Opt., 19, 281
Rutledge, G. K., Alpert, J., & Ebisaki, W. 2006, Bull. Am. Meteorol. Soc., 87, 327
Sela, J. 1980, Mon. Weather Rev., 108, 1279
Schöck, M., et al. 2009, PASP, 121, 384
Tatarski, V. I. 1961, Wave Propagation in a Turbulent Medium (New York: Dover)
Tokovinin, A. 2002, PASP, 114, 1156
Trinquet, H., & Vernin, J. 2006, PASP, 118, 756
Walker, M. F. 1970, PASP, 82, 672
Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., & Toth, Z. 2008, Mon. Weather Rev., 136, 463
Xu, K.-M., & Randall, D. A. 1996, J. Atmos. Sci., 53, 3084
Yang, F., Pan, H.-L., Moorhi, S., Lord, S., & Krueger, S. 2005, in Proc. Fifteenth ARM Science Team Mtg., 1, http://www.arm.gov/publications/proceedings/conf15/extended_abs/yang_ f.pdf

Ye, Q.-Z., & Chen, S.-S. 2010, J. Atmos. Sci., submitted