Online Supplemental Material

An analogy-based method for strong convection forecasts in China using GFS forecast data

Na LI\textsuperscript{a}, Lingkun RAN\textsuperscript{a} and Baofeng JIAO\textsuperscript{b}

\textsuperscript{a}Key Laboratory of Cloud-Precipitation Physics and Severe Storms, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China; \textsuperscript{b}Nanjing University of Information Science and Technology, Nanjing, China

Text S1 Evaluation parameters for the analogy-based strong convection forecast method

1. Equitable threat score (ETS)

The ETS describes the skill of the method in terms of having a high hit rate and low false alarm rate. Following Table S1, the ETS is defined as

\[ ETS = \frac{N_a - R_a}{N_a + N_b + N_c - R_a}, \] (S1)

where

\[ R_a = \frac{(N_a + N_b)(N_a + N_c)}{N_a + N_b + N_c + N_d}. \] (S2)

Perfect forecasts have an ETS of 1; random or constant forecasts have an ETS of 0.

2. Critical success index (CSI)

Like the ETS, the CSI also describes the skill of the method in terms of having a
high hit rate and low false alarm rate. The CSI is defined as

$$\text{CSI} = \frac{N_a}{N_a + N_b + N_c}.$$  \hspace{1cm} (S3)

Perfect forecasts have a CSI of 1; no hits means a CSI of 0.

3. Probability of detection (POD)

POD describes the skill of the method in forecasting the occurrence of strong convections, and is defined as

$$\text{POD} = \frac{N_a}{N_a + N_b}.$$  \hspace{1cm} (S4)

If the strong convections are all predicted, POD = 1; if none is predicted, POD = 0.

4. False alarm ratio (FAR)

The FAR represents the proportion of forecasts of strong convection occurrence that do not actually happen:

$$\text{FAR} = \frac{N_c}{N_a + N_c}.$$  \hspace{1cm} (S5)

A FAR of 0 (1) indicates that strong convection occurrences are all (never) predicted.

5. Brier Score (BS)

The BS describes the skill of the method in predicting the occurrence probabilities of strong convections:

$$\text{BS} = \frac{1}{N} \sum_{i=1}^{N} (P_{fi} - P_{oi})^2,$$  \hspace{1cm} (S6)

where $P_{fi}$ is the occurrence probability of the strong convection at time $i$, and $P_{oi}$ is the observed occurrence probability of the strong convection at the same time ($P_{oi} = 1$ when the strong convection happens; $P_{oi} = 0$ when no strong convection happens). A BS of 0 (1) indicates a perfect probability forecast of strong convection occurrence (non-occurrence).
The analogy-based method for thunderstorm forecast calculates 17 convective parameters which is defined as follows.

1. PW (precipitable water)

The depth of liquid water at the surface that would result after precipitating all of the water vapor in a vertical column over a unit area (unit: cm):

$$ PW = \frac{1}{g \rho_w} \int_{p_0}^{p_t} q dp, \quad (S7) $$

where $q$ is specific humidity, $p_0$ is surface pressure, $p_t$ is pressure at the top-level of the data (100 hPa here), $g$ is gravitational parameter, $\rho_w$ is water density ($\rho_w = 1.0 \ g \ cm^{-3}$).

2. LI (lifting index)

The temperature difference between the environment $T_e$ and an air parcel lifted adiabatically $T_p$ at a given pressure height in the troposphere of the atmosphere, usually 500 hPa (unit: °C), namely

$$ LI = T_e - T_p. \quad (S8) $$

When $LI<0$, the atmosphere is unstable.

3. CAPE (convective available potential energy)

CAPE represents the amount of buoyant energy available to accelerate a parcel vertically, or the amount of work a parcel does on the environment. CAPE is the positive area on a sounding between the parcel's assumed ascent along a moist adiabat and the environmental temperature curve from the level of free convection (LFC) to
the equilibrium level (EL). The greater the temperature difference between the warmer parcel and the cooler environment, the greater the CAPE and updraft acceleration to produce strong

\[
\text{CAPE} = \int_{p_{\text{EL}}}^{p_{\text{LFC}}} R_d (T_{vp} - T_{ve}) \ln \frac{dp}{T} = - \int_{p_{\text{LFC}}}^{p_{\text{EL}}} R_d (T_{vp} - T_{ve}) \ln \frac{dp}{T}
\]  

(S9)

The "\[\int\]" symbol here represents a vertical integration between the LFC and the EL, \(T_{vp}\) and \(T_{ve}\) are respectively the virtual temperatures of the parcel and the environment.

4. CIN (convection inhibition energy)

CIN represents the amount of negative buoyant energy available to inhibit or suppress upward vertical acceleration, or the amount of work the environment must do on the parcel to raise the parcel to its LFC,

\[
\text{CIN} = \int_{p_{\text{LFC}}}^{p_{\text{EL}}} R_d (T_{ve} - T_{vp}) \ln \frac{dp}{T}.
\]

(S10)

5. \(K\) index

The \(K\) Index is a linear combination of temperature and dewpoint at various levels, namely

\[
K = \frac{1}{2} (T_0 + T_{850}) - T_{500} + \frac{1}{2} (T_{d0} + T_{d850}) - (T - T_d)_{700},
\]

(S11)

where, \(T\) and \(T_d\) are respectively temperature and dew-point temperature; the subscripts "0", "850", "500" denote the surface, 850 hPa, and 500 hPa, respectively. The unit is °C. The \(K\) index includes the temperature lapse rate, the low-level moisture condition, and the mid-level saturation degree. The larger of \(K\) index, the
more favorable for the thunderstorm occurrence.

6. IC (convective stability index)

IC reflects the static instability of mid troposphere, which is defined as

$$IC = \theta_{e500} - \theta_{e850}$$  \hspace{1cm} (S12)

where, $\theta_e$ is equivalent potential temperature, the subscripts “850”, “500” denote 850 hPa and 500 hPa, respectively.

7. IL (conditional instability index)

IL reflects the conditional instability of the atmosphere, which is defined as

$$IL = \theta_{se500} - \theta_{se0}$$  \hspace{1cm} (S13)

where, $\theta_{se}$ is saturated equivalent potential temperature, the subscripts “0”, “850” denote surface and 850 hPa, respectively.

8. ILC (conditional convective instability index)

ILC is the sum of IL and IC, which is defined as,

$$ILC = (\theta_{se500} - \theta_{se0}) + (\theta_{e500} - \theta_{e850})$$  \hspace{1cm} (S14)

9. DCI (deep convective index)

DCI combines the properties of equivalent potential temperature at 850 mb with instability (unit: °C),

$$DCI = T_{850} + T_{d850} - LI$$  \hspace{1cm} (S15)

DCI values of roughly 30 or higher indicate the potential for strong thunderstorms.

10. MDPI (microburst day potential index)

MDPI is used to determine whether the microburst is happened or not,
\[ MDPI = \frac{\theta_{e_{\text{max}}} - \theta_{e_{\text{min}}}}{CT} \]  
(S16)

where, \( \theta_{e_{\text{max}}} \) is the maximum equivalent potential temperature from the surface layer to 500-hPa height, \( \theta_{e_{\text{min}}} \) is the minimum equivalent potential temperature from 650 hPa to 500 hPa, the unit is °C, CT is critical values (CT = 20°C).

11. TT (total temperature)

TT consists of two components, the lapse rate between 850 and 500 mb and the 850 mb dewpoint,

\[ TT = T_{850} + T_{d_{850}} - 2T_{500}, \]  
(S17)

which accounts for both static stability and 850 mb moisture.

12. SWEAT (severe weather threat index)

SWEAT evaluates the potential for severe weather by combining several parameters into one index. These parameters include low-level moisture (850 hPa dewpoint), instability (TT), lower and middle-level (850 and 500 hPa) wind speeds, and warm air advection (veering between 850 and 500 mb),

\[ \text{SWEAT} = 12T_{d_{850}} + 20(\text{TT} - 49) + 2f_{850} + f_{500} + 125(S + 0.2) \]  
(S18)

Where, \( f_{850} \) is the wind speed at 850 hPa, \( f_{500} \) is the wind speed at 500 hPa, \( S = \sin(\alpha_{500} - \alpha_{850}) \), \( \alpha_{500} \) and \( \alpha_{850} \) is the wind directions of 500 hPa and 850 hPa winds.

13. EHI (energy helicity index)

EHI attempts to combine CAPE and storm-relative helicity \( H_{SR} \) into one index to assess the potential for supercell and mesocyclone development,

\[ \text{EHI} = \frac{(H_{SR} \cdot \text{CAPE})}{160000} \]  
(S19)

High EHI values represent an environment possessing high CAPE and/or high S-R helicity.
14. SWISS (stability and wind shear index for thunderstorms in Switzerland)

SWISS is used to forecast the occurrence of thunderstorm, which is defined as,

\[
\text{SWISS} = \text{LI} + 0.3\text{Shr}_{0.3} + 0.3(T - T_d)_{650}
\]  

(S20)

Where, Shr$_{0.3}$ is the density-weighted mean vertical wind shear.

15. WINDEX (wind index)

WINDEX represents a potential of the surface damaging wind influence by the mid-level and low-level temperature and humidity,

\[
\text{WINDEX} = 5[H_m R_Q (\Gamma^2 - 30 + Q_L - 2Q_m)]^{0.5},
\]  

(S21)

where $H_m$ is the height of the 0°C level, $R_Q = \frac{Q_L}{12}$ which should be less than 1 g kg$^{-1}$, $\Gamma$ is the temperature rate between the surface and the 0°C level, $Q_L$ is the averaged mixing ratio below 1-km height, $Q_m$ is water vapor mixing ratio at 0°C level.

16. BRN (bulk Richardson number)

BRN usually is a decent indicator of convective storm type within given environments. It incorporates buoyant energy (CAPE) and the vertical shear of the horizontal wind, both of which are critical factors in determining storm development, evolution, and organization.

\[
\text{BRN} = \frac{\text{CAPE}}{0.5V_{shr}^2},
\]  

(S22)

where

\[
V_{shr} = \left(\int_{0}^{6000} \rho_z |V_z| dz - \frac{1}{2} \left[V_0 + V_{500}\right] \right) / 6000,
\]  

(S23)

The subscript “z” denote the height, “0” is the surface, “500” and “6000” are the heights with the unit “m”.
17. SSI (storm strength index)

SSI combines CAPE and vertical wind shear in the low atmosphere to distinguish the strong thunderstorm and ordinary thunderstorm,

\[
SSI = 100[2 + (0.276 \ln(Shr)) + (2.011 \times 10^{-4} \text{CAPE})], \quad (S24)
\]

where \( Shr \) is the density-weighted mean vertical wind shear below 3.6 km height.

| Forecast                                      | Observed strong convection times | Observed non-strong-convection times | Total   |
|-----------------------------------------------|----------------------------------|-------------------------------------|---------|
| Strong convection times                       | \( N_a \)                        | \( N_c \)                            | \( N_a+N_c \) |
| Non-strong convection times                   | \( N_b \)                        | \( N_d \)                            | \( N_b+N_d \) |
| Total                                         | \( N_a+N_b \)                    | \( N_c+N_d \)                        | \( N \)   |
**Figure S1** Process through which the analogy-based strong convection forecasts are created.