Empirical and Statistical Forecast Guidance

Publish date: 7/20/2016 (updates will be posted as required)
Information prepared by: Dan Harnos, Peitao Peng and Jon Gottschalck

In preparation of the experimental Week 3-4 temperature and precipitation outlooks, the CPC currently also utilizes statistical forecast tools which are in various levels of development to complement dynamical model guidance.

1. ENSO-MJO Phase Model

Information prepared by Daniel Harnos (CPC)

(a) Input Data (December 1979-January 2014)

- Temperature data: CPC Unified land only T_{2m} data
  - Originally T_{max} and T_{min} at 0.5° resolution.
  - Converted to mean temperature and bilinearly interpolated to 1x1
- Precipitation data: CPC Unified Gauge-based
  - Originally 7 day running accumulation at 1° resolution.
  - Converted to ¼ power to increase normality of distribution.
- ENSO data: CPC Oceanic Nino Index (ONI) found at http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml
  - Dataset used was developed from ERSST.v3b
  - Traditional divisions are utilized for El Niño (≥0.5) and La Niña (≤-0.5) with an additional ENSO neutral class between the two.
- MJO data: Wheeler and Hendon (2004) Real-time Multivariate MJO (RMM) Index found at http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt
  - Eight MJO states are used in the traditional sense, where the amplitude of the two RMMs must be ≥1 with an additional weak MJO class when amplitude is <1.

(b) Methodology

The underlying assumption of the phase model is that ENSO, MJO, and trend influences are discrete from one another and additive. First, anomalies for each grid cell are pooled for 3 month running periods for T_{2m} and P. The median values are then taken for each grid cell, to enable classification of whether values are above/below normal. The initial state of ENSO (3 classes) and MJO (9 classes) are taken. Conditional distributions of T_{2m} and P are then calculated for future periods of Week 3 (days 15-21), Week 4 (days 22-28), and a combined Weeks 3+4 (days 15-28) relative to each of the 3 ENSO and 9 MJO base states. Conditional means and variances of T_{2m} and P are taken for each of the 3 ENSO and 9 MJO base states. ENSO and MJO impacts are assumed to be independent of one another, thus allowing the means and variances for ENSO and MJO to be summed together. The mean of the combined

ENSO and MJO distribution is also shifted to account for the long-term trend (30 y) that is assumed to be linear.

A Gaussian distribution of T_{2m} and P can then be constructed utilizing the sum of the three means (ENSO, MJO, and long-term trend) and the sum of the two variances (ENSO and MJO). Note due to the P distribution being non-Gaussian, a ¼ power transformation is applied to the anomaly values as specified above. The forecast distributions of T_{2m} and P for the current base states of ENSO, MJO, and point in time relative to the long-term trend can then be compared to the climatological distribution to quantify any shift in the probabilities. An example of the methodology is shown below.

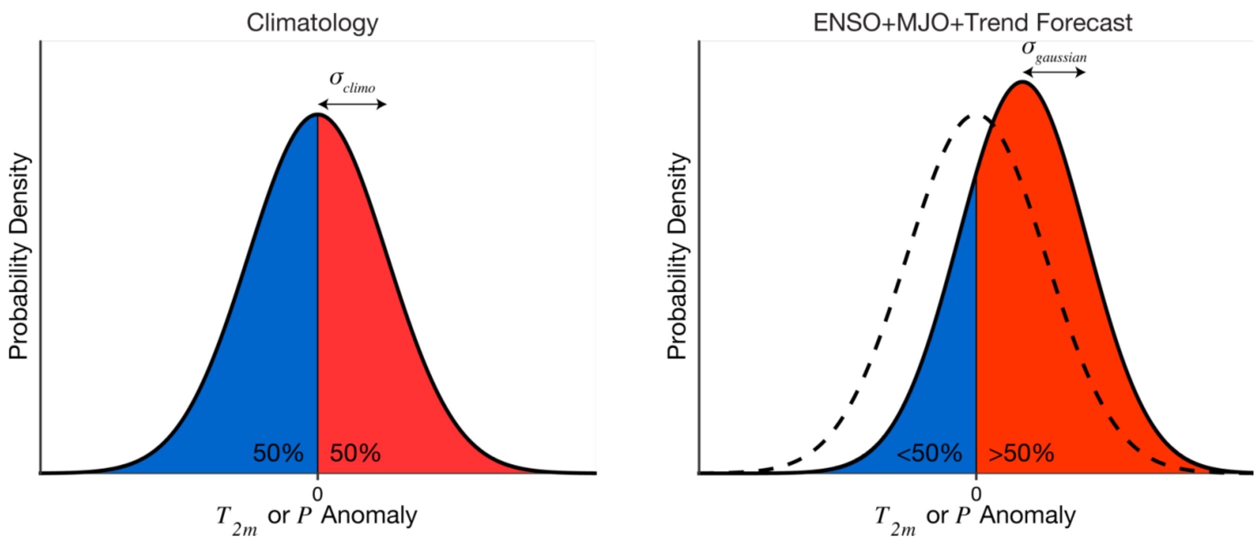

Figure 1: Schematic of forecast calculations for each grid point by the ENSO-MJO phase model. The climatological distribution is calculated, with the area below the curve shaded in blue (red) denoting below (above) normal. Note the climatological distribution is shown as Gaussian, but is non-parametric in practice. For each ENSO and MJO base state an adjusted forecast PDF (right) that is assumed to be Gaussian is generated with the distribution mean shifted for ENSO and MJO subsets and additional long-term trend while the distribution variance is shifted for combined ENSO and MJO subsets. The forecast distribution is compared to the climatological distribution (dashed line), with proportion of area residing under the forecast distribution's curve compared to the climatological median permitting adjusted probabilistic forecasts of either T_{2m} or P.

Heidke Skill Scores provided for the tool are calculated using a leave-one-out method of cross-validation. For more details on the tool see Johnson et al. (2014).

- Caveats
  - Modulation of ENSO/MJO impacts on one another are assumed to be linear and independent of one another.
  - Long-term trend is assumed to be linear.
  - Distributions of T_{2m} and P are assumed Gaussian.
  - No information about ENSO and MJO strength is conveyed.

- o Transitions can be abrupt relative to the RMM magnitudes when the MJO signal is near an amplitude of one or one of the boundaries between the 8 phases. Caution should be used in such situations.
- o Week 3 and Week 4 individually do not always sum to Weeks 3+4 in this framework. Week 3 experiences large sample sizes than the other two periods, due to the code being trained on when the full analysis period had to reside within the running 3-month period the base state was in. For instance, with a base state of JFM March 25-31 spanning week 3 and April 1-7 spanning week 4 the former would be considered while the latter would be neglected (as would the combined week 3-4 period).

- References

- o Johnson, N. C., D. C. Collins, S. B. Feldstein, M. L. L'Heureux, and E. E. Riddle, 2014: Skillful wintertime North American temperature forecasts out to 4 weeks based on the state of ENSO and the MJO. *Wea. Forecasting*, **29,** 23-38.
- o Wheeler, M. C. and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132,** 1917-1932.

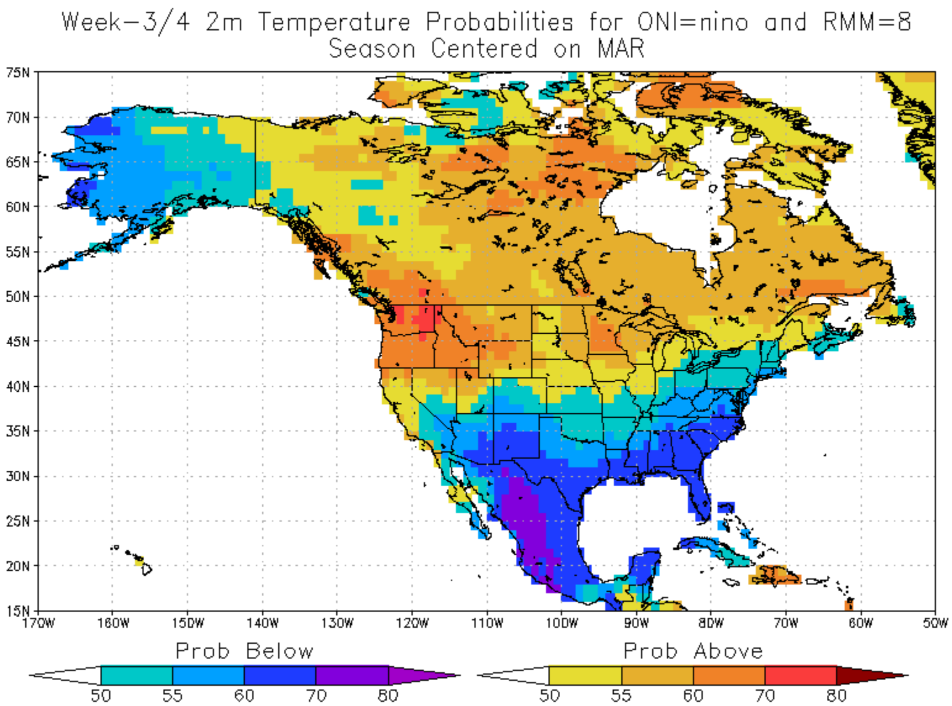

Figure 2: Example forecast map made March 4, 2016 which indicates temperature probabilities for the above and below normal category for a case of El Nino conditions along with the MJO located in Phase 8 initialized during March.

2. <u>Multiple Linear Regression Model</u>
Information prepared by Daniel Harnos (CPC)

(a) Input Data (January 1982-February 2014)

- Temperature data: Wei Shi's internal CPC T_{2m} data found at `/cpc/sfc_temp/ARCHIVE/GRID/`
  - Originally T_{max} and T_{min} at 0.5° resolution.
  - Converted into a mean temperature and bilinearly interpolated to 1° resolution.
  - Predictor of standardized linear trend over the 1982-2014 period.
- Precipitation data: CPC Unified Gauge-based data found at `/cpc/home/aallgood/data/observations/land_air/all_ranges/global/global_precipitation/gridded/1.0DEG/binary/`
  - Originally 7 day running accumulation at 1° resolution.
  - Converted to ¼ power to increase normality of distribution.
  - Predictor of standardized linear trend over the 1982-2014 period.
- ENSO data: Niño 3.4 from OISSTv2 ftp://eclipse.ncdc.noaa.gov/pub/OI-daily-v2/NetCDF/
  - One predictor, most recent 2-week running mean anomaly (standardized).
  - First four harmonics removed as the seasonal cycle.
- MJO data: CPC Real-time Multivariate MJO (RMM) Index used in generating the analysis seen at http://www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/whindex.shtml
  - Two predictors, most recent daily RMM1 and RMM2. These are orthogonal to one another, and uncorrelated with the Niño 3.4 anomaly.
  - Linear trend removed from each RMM index.

(b) Methodology

The underlying assumption of the phase model is that ENSO, MJO, and long-term trend influences are discrete from one another and additive. First, anomalies for each grid cell are pooled for 3 month running periods for T_{2m} and P. The median values are then taken for each grid cell, to enable classification of whether values are above/below normal. Conditional distributions of T_{2m} and P are then calculated for the combined future period of Weeks 3+4 (days 16-29, as opposed to days 15-28 due to the ENSO and MJO information having a latency of 1 day). Regression coefficients are generated over the training period based on the observed future T2m and P anomaly relationships to the initial RMM1, RMM2, Niño 3.4, and linear trend predictors. The regression coefficients developed from the training period can then be applied to real-time data to generate forecast anomalies from each predictor in support of forecast operations.

Gaussian distributions of forecast T_{2m} and P are then constructed utilizing the sum of the four forecast mean anomalies (Nino 3.4, RMM1, RMM2 and long-term trend) along with the climatological variance. The variance of the Gaussian forecast distribution is then adjusted based upon how much variance was explained by the regression tool over the training period (r-squared adjustment) in order to compensate for overconfidence in the forecasts and yield more reliable probabilities. Note due to the P distribution being non-Gaussian, a ¼ power transformation is applied to the anomaly values as specified above. The forecast distributions of T_{2m} and P for the current base states of ENSO, MJO, long-term trend can then be compared to the climatological distribution to quantify any shift in the probabilities (similar to the probability distribution function graphic accompanying the phase model). In addition to the probabilistic guidance provided to the Week 3-4 forecasters, forecast anomalies are provided from: the combined 4 predictors, ENSO, MJO (RMM1 and RMM2 combined influences) and long-term trend. The forecast anomalies from each component help provide the forecaster with information as to which sources of climate variability are influencing the forecast distribution and to what relative extent.

Heidke Skill Scores provided for the tool are calculated using a leave-one-out method of cross-validation. A formal publication is forthcoming on the tool. For comparison with the ENSO-MJO phase model, Heidke Skill Scores are also provided based on the background ONI state (El Nino, Neutral, La Nina) and MJO state (Phases 1-8 or Weak MJO). Retrospective skill analyses over the training period reveals the Phase Model and Regression contain independent information from one another, suggesting for certain situations climate variability magnitude matters while in others it does not.

- Caveats
    - Linearity is assumed among ENSO/MJO/trend responses.
    - Long-term trend is assumed to be linear.
    - Distributions of T_{2m} and (adjusted) P are assumed Gaussian. The tool was also evaluated using an empirical distribution framework, revealing this assumption to be generally valid.

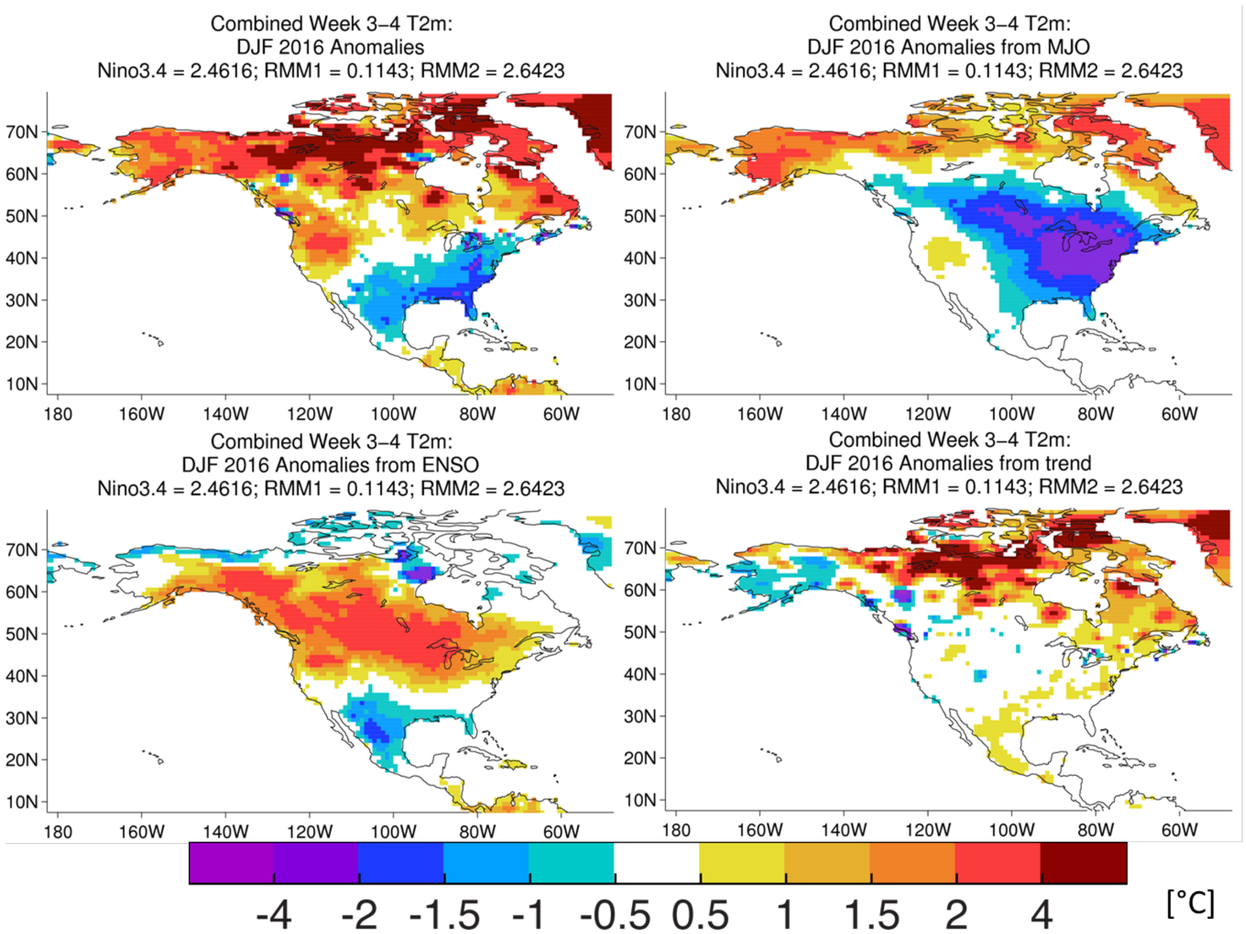

Figure 3: Example statistical forecast guidance from the MLR tool. Temperature anomalies in degrees C for combined contributions of ENSO, MJO and trend (top left), ENSO only (bottom left), MJO only (top right) and trend only (bottom right).

3. Constructed Analogue

Information prepared by Peitao Peng (CPC)

(a) Input Data (Jan 1979-current)

1) 200hPa stream function (S_{200}) data: CDAS at /cpc/analysis/cdas/daily/prs

2) Surface air temperature (T_{2m}) data: Wei Shi's internal CPC T_{2m} data at `/cpc/sfc_temp/ARCHIVE/GRID/`

3) Precipitation (P) data: CPC Unified Gauge-based data found at /cpc/prcp/PRODUCTS/CPC_UNI_PRCP/GAUGE_GLB/CTLPRCP_CU_GAUGE_V1.0GLB_0.50deg.lnx.ctl

4) 500hPa geopotential height (Z_{500}) data: CDAS at /cpc/analysis/cdas/daily/prs

Weekly mean data are calculated in their original resolution, and then interpolated to 2.5^{o} x 2.5^{o} resolution. In order for the precipitation data to be closer to normal distribution, ¼-power was applied.

(b) Methodology

Constructed analogue (CA) method first constructs an analogue of an initial condition of a predictor with a weighted average of historical data, and then constructs a forecast by applying the same weights to the lagged predictand data in history. The weights are obtained by minimizing the root-mean-square error of the constructed analogue of the initial condition. In week3-4 forecast, the predictor is chosen to be the weekly mean S_{200} over the tropics and northern hemisphere, and the predictands are weekly mean T_{2m}, P and Z_{500}. All the input data to the CA model are linearly detrended. The final forecast is the linear combination of the week3-4 mean of the CA forecast with the linear trend of the corresponding variables.

In order to convert the deterministic forecast (e.g., the forecasted anomalies) into a 2-class probabilistic format, the probability distribution function (PDF) of each variable is determined with their standard deviation calculated form historical data and the assumption that these data are in normal distribution. The probability for a forecasted anomaly to happen is calculated by locating the anomaly in the PDF.

Heidke Skill Scores provided for the tool are calculated using a 3-year-out cross-validation.

- **References**

van den Dool, H.M., 1994: Searching for analogues, how long must we wait? Tellus, 46A, 314-324.

van den Dool, H. M., 2007: Empirical Methods in Short-Term Climate Prediction. Oxford University Press, 215 pp.

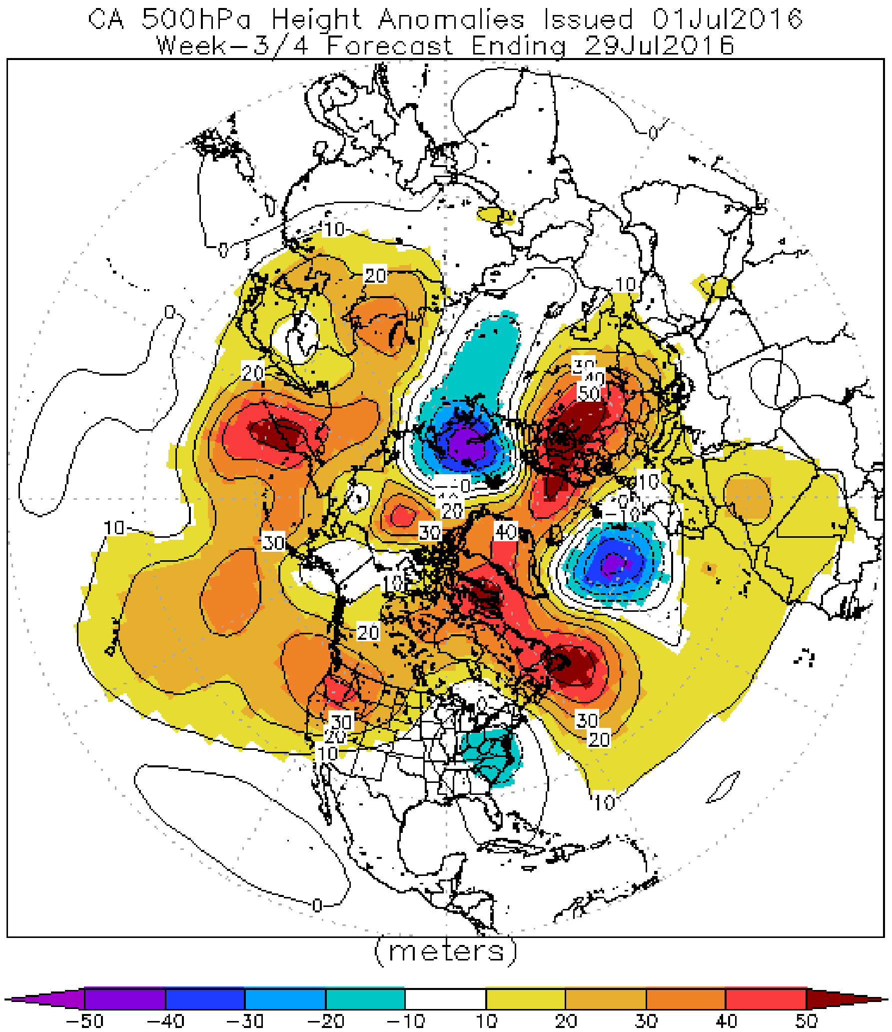

Figure 4: Example constructed analogue 500-hPa height anomaly (m) forecast from July 1, 2016. Positive height departures in yellow/red colors and negative height departures expressed in blue/purple shades.

4. <u>Coupled Linear Inverse Model (C-LIM)</u>
Information prepared by Jon Gottschalck (CPC)

Information from this guidance is forecast precipitation and winds for the Tropics only. We worked with our partners at the Earth System Research Laboratory (ESRL) to implement operationally at CPC the code developed by ESRL staff. The purpose of this guidance is to have a complement to the subseasonal forecasts of tropical convection and winds to that provided from the dynamical models. The forecast tool produces pentad forecasts of Outgoing Longwave Radiation (OLR) and 200-hPa vector winds among other variables. The forecasts are updated every 5 days. The reference below describes the methodology and other information in considerable detail.

Newman, M., P. D. Sardeshmukh, and C. Penland, 2009: How important is air-sea coupling in ENSO and MJO evolution? *J. Climate*, 22, 2958-2977.

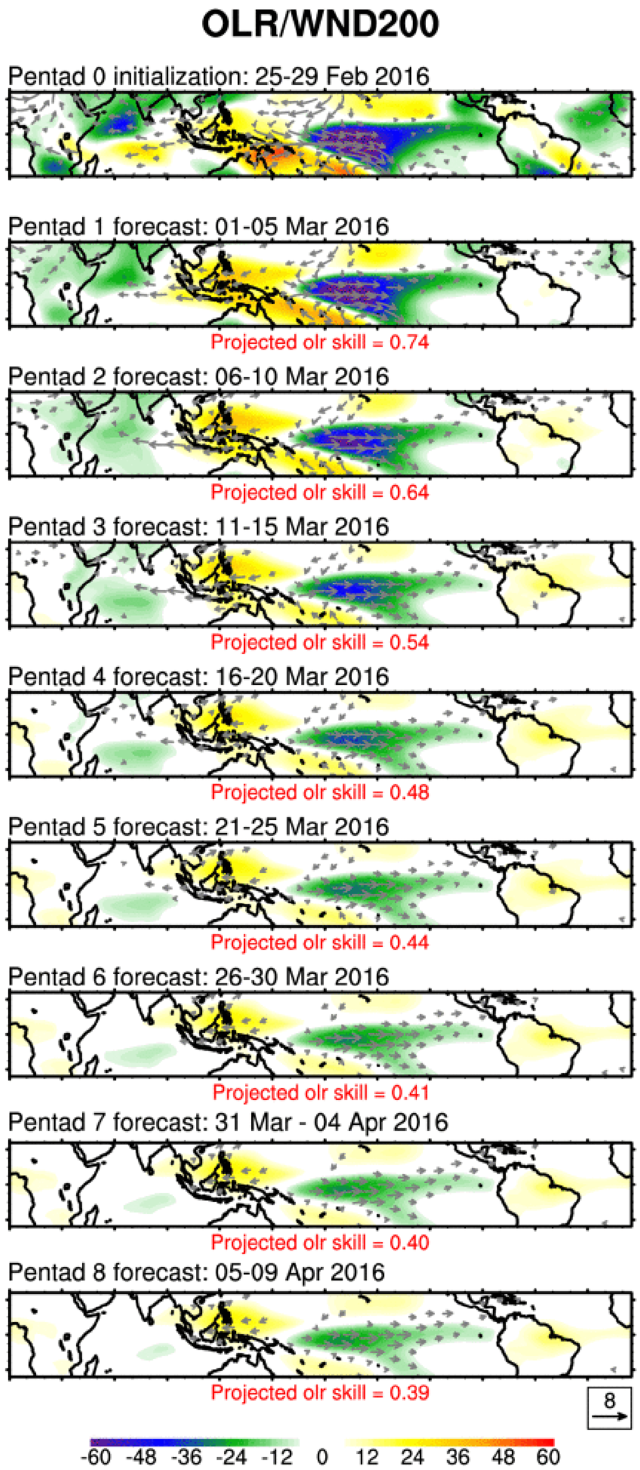

Figure 5: Example forecast map from the C-LIM tool made during late February 2016. Shown are successive pentad forecasts of OLR anomalies (shading) and 200-hPa vector wind in the Tropics.