From Probability Forecast to Probability Distribution Forecast Evaluation

Applying Bradley-Schwartz summary measures to evaluate ensemble forecast model outputs

教育訓練

計畫: 105 年度全球預報模式之2 至4週統計後處理系統發展
地點: 中央氣象局
時間: 10月31日 3:00 PM – 5:00 PM

李彥緯 博士

Outline

- Deterministic categorical forecasts
- Deterministic continuous forecasts
- Probability forecast for Ensemble forecasts system
- Evaluation measures of probability distribution forecast
- Application in ECMWF S2S data

Forecasters Questions

- How can the forecasting system be improved?
- How useful are these forecast products?
- Has the forecasting performance of our institution improved?

What is forecast verification?

- Forecast verification is the exploration and assessment of the **quality** of a forecasting system based on a sample or samples of previous forecasts and corresponding observations.

- What is meant by quality?

  The degree of correspondence between forecasts and observations

Types of Forecasts

- **Categorical**

  Discrete variables, a finite set of predefined values.

– **Ordinal**: the categories provide a ranking of the data

  $(YES\colon\{>,<\}; NO\colon\{+,-,\times,\div\})$

  {cold, mild, hot}

– **Nominal**: no natural ordering of the categories

  $(NO\colon\{>,<,+,-,\times,\div\})$

  {0, 1} ,{Yes, No}, {Rain, No Rain}

- **Continuous**

  Variables such as mean sea level pressure (MSLP) or temperature over a region.

  $(YES\colon\{>,<,+,-,\times,\div\})$

Types of Forecasts

- **Deterministic:**
  a specific category or particular value for either a discrete or continuous variable
  - a class attribution, e.g. there will be rain tomorrow
  - a single number, e.g. the temperature tomorrow

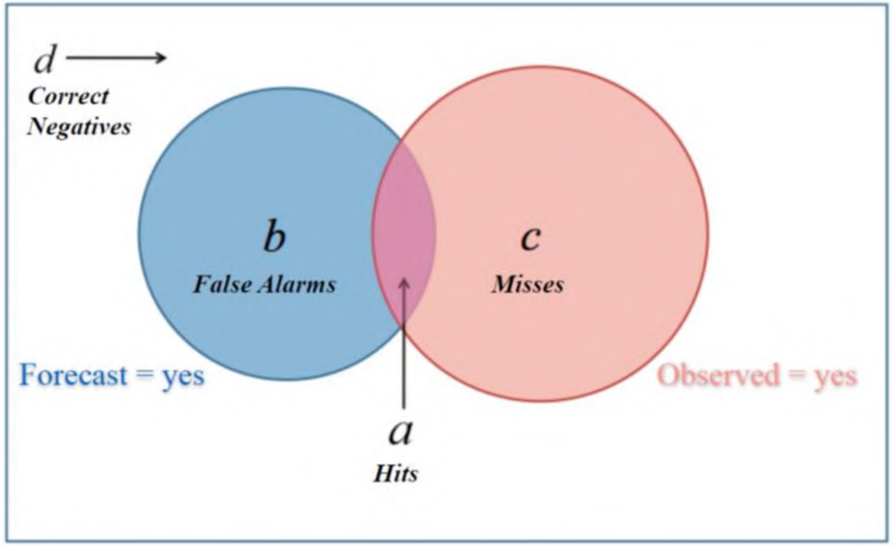
- **Probabilistic:**
  expresses the degree of forecast uncertainty
  - a probability, e.g. P(rainfall tomorrow) = 0.2
  - a pdf, e.g. a distribution for temperatures tomorrow

Deterministic Categorical Forecasts

- **Binary Forecasts**

  Y = {yes, no}, {1, 0}

  e.g. it will or will not rain tomorrow



- **Contingency Table**

Forecast	Observed		
	Yes	No	Total
Yes	a	b	$a+b$
No	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d=n$

Event forecast	Event observed	
	Yes	No
Yes	Hit	False alarm
No	Miss	Correct rejection

Simple Scores

		Observation		
		yes	no	
Forecast	yes	a hits	b false alarms	$a+b$ yes fcsts
	no	c misses	d correct rejects	$c+d$ no fcsts
		$a+c$ yes obs	$b+d$ no obs	N total fcsts
		Marginal of Obs		Marginal of Fcst

- **Bias Score**:

$$B = \frac{a+b}{a+c} = \frac{\#\ \text{forecasted events}}{\#\ \text{observed events}}$$

  - B = 1 **unbiased** (perfect forecast),
  - B < 1 underforecast, B > 1 overforecast

  Bias alone conveys no information about skill, because any value can be attained by changing the decision threshold.

Fore-cast	Tornado observed: Yes	No	Total
Yes	2	98	100
No	49	2654	2703
Total	51	2752	2803

Fore-cast	Tornado observed: Yes	No	Total
Yes	14	37	51
No	37	2715	2752
Total	51	2752	2803

Simple Scores

Forecast \ Observation	yes	no	Marginal of Fcst
yes	a hits	b false alarms	$a+b$ yes fcsts
no	c misses	d correct rejects	$c+d$ no fcsts
Marginal of Obs	$a+c$ yes obs	$b+d$ no obs	N total fcsts

- **Hit Rate (Probability of Detection)** :
  The proportion of occurrences that were correctly forecast

$$H = \frac{a}{a+c} = \frac{\#\, hits}{\#\, observed\ events}$$

  - $0 \leq H \leq 1$, best score: $H = 1$, best score ≠ perfect forecast

  A threshold probability of 0, meaning that occurrence is always forecast, gives $H = 1$, and a threshold probability of 1, meaning that the event is never forecast, gives $H = 0$.

Simple Scores

		Observation: yes	Observation: no	Marginal of Fcst
Forecast	yes	a hits	b false alarms	$a+b$ yes fcsts
	no	c misses	d correct rejects	$c+d$ no fcsts
Marginal of Obs		$a+c$ yes obs	$b+d$ no obs	N total fcsts

- **False Alarm Ratio**:
  A estimate of the conditional probability of a false alarm given that occurrence was forecast.

$$FAR = \frac{b}{a+b} = \frac{\#\,false\ alarms}{\#\,forecasted\ events}$$

  - $0 \le FAR \le 1$, best score: $FAR = 0$, best score ≠ perfect forecast

- **False Alarm Rate (Probability of False Detection)** :

$$F = \frac{b}{b+d} = \frac{\#\,false\ alarms}{\#\,non\ events}$$

  - $0 \le F \le 1$, best score: $F = 0$, best score ≠ perfect forecast
  - F is analogous to the probability of a **Type I error**

Medical Statistics

		Facts	
		Positive	*Negative*	
Prediction	*Positive*	True Positive	False Positive (Type I error, P-value)	→ Positive predictive value
	Negative	False Negative (Type II error)	True Negative	→ Negative predictive value
		↓ Sensitivity	↓ Specificity	

- **Sensitivity** (Hit Rate) (= 1 – Type II Error)= $\frac{a}{a+c}$

- **Specificity** (=1 – Type I Error) = $\frac{d}{b+d}$ (=1 – False Alarm Rate)

- **Positive predictive value** (=1 – False Alarm Ratio) = $\frac{a}{a+b}$

- **Negative predictive value** = $\frac{d}{c+d}$

Forecast	Observed		
	Yes	No	Total
Yes	a	b	$a+b$
No	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d=n$

Simple Scores

		Observation: yes	Observation: no	Marginal of Fcst
Forecast	yes	a hits	b false alarms	$a+b$ yes fcsts
	no	c misses	d correct rejects	$c+d$ no fcsts
Marginal of Obs		$a+c$ yes obs	$b+d$ no obs	N total fcsts

- **Proportion Correct** (Accuracy):

$$PC = \frac{a+d}{N} = \frac{\#\ correct\ forecasts}{\#\ forecasts}$$

  – Proportion of correct forecasts
  – 0 ≤ PC ≤ 1, best score: PC = 1, best score = perfect forecast

The optimal threshold probability that maximises PC is 0.5, and hence the PC score can always be maximised by forecasting occurrence of the event whenever the observed probability of the event exceeds 0.5.

Limitations of Simple Scores

- How large is a good score?
- Best score not necessarily perfect forecast
- Hedging (Playing) a score

Generic Form of a Skill Score

- Relative measure of the quality of the forecasting system compared to some (usually low-skill) benchmark forecast.

$$SS = \frac{A - A_{ref}}{A_{perf} - A_{ref}}$$

A: accuracy score (e.g. *PC*)

A*ref*: accuracy of reference forecast (e.g. random)

A*perf*: accuracy of perfect forecast (the best possible score)

– SS = 1 perfect forecast

– SS > 0 skillful, better than reference;

– SS < 0 less skillful than reference

Heidke Skill Score

Forecast	Observed		
	Yes	No	Total
Yes	a	b	$a+b$
No	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d=n$

- Generic Skill Score with A = *PC* and **reference = random forecast**

$$A = \left(\frac{a+d}{N}\right) \qquad A_{perf} = 1$$

$$A_{ref} = \left(\frac{(a+b)}{N}\right)\cdot\left(\frac{(a+c)}{N}\right) + \left(\frac{(d+c)}{N}\right)\cdot\left(\frac{(d+b)}{N}\right)$$

- Heidke Skill Score

  based on Accuracy corrected by the number of hits that would be expected by chance.

$$HSS = \frac{ad - bc}{\left((a+c)\cdot(c+d) + (a+b)\cdot(b+d)\right)/2}$$

$$-\infty < HSS \le 1, \quad HSS \le 0 \ \text{ no skill}$$

Deterministic Continuous Forecasts

- Sample, forecast-observation pairs (real valued)

$$\{y_i, o_i\}, \qquad i = 1 \dots N$$

- **Sample Means**

$$\bar{y} = \frac{1}{N}\sum_i y_i, \quad \bar{o} = \frac{1}{N}\sum_i o_i$$

- **Sample Variance**

$$s_y^2 = \frac{1}{N}\sum_i \left(y_i - \bar{y}\right)^2, \quad s_o^2 = \frac{1}{N}\sum_i \left(o_i - \bar{o}\right)^2$$

Simple Error Scores

- **Mean Error** (Bias):

$$B = \bar{y} - \bar{o}$$

- **Mean Absolute Error**:

$$MAE = \frac{1}{N}\sum_i \left| y_i - o_i \right|$$

- **Mean Squared Error** (MSE), **Root MSE** (RMSE):

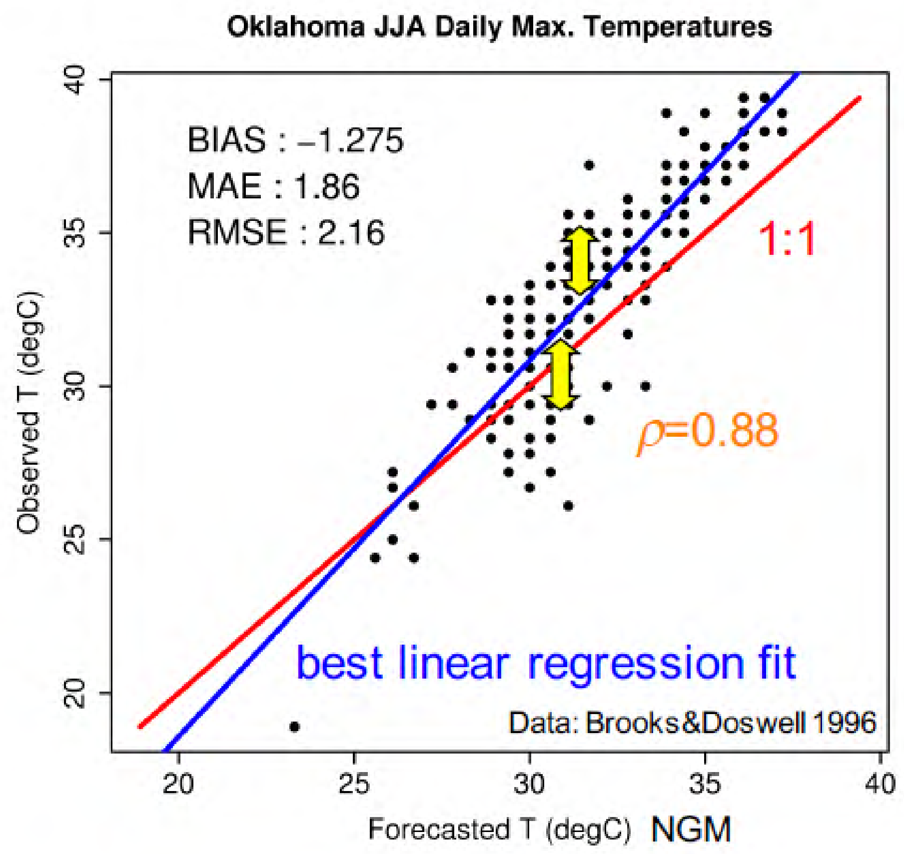
$$MSE = \frac{1}{N}\sum_i (y_i - o_i)^2, \quad RMSE = \sqrt{MSE}$$

Sensitive to outliers, dominated by large deviations

Correlation Coefficient

- Linear Correlation Coefficient

$$\rho = \frac{\frac{1}{N}\sum_{i}^{N}(y_i - \overline{y})\cdot(o_i - \overline{o})}{s_y \cdot s_o}$$



  - $-1 \le \rho \le 1$, $\rho = 1$ best score
  - A measure of random error (scatter around best fit)
  - ρ^2: fraction of variance in observations explained by best linear model
  - ρ measures potential skill

Regression Slope

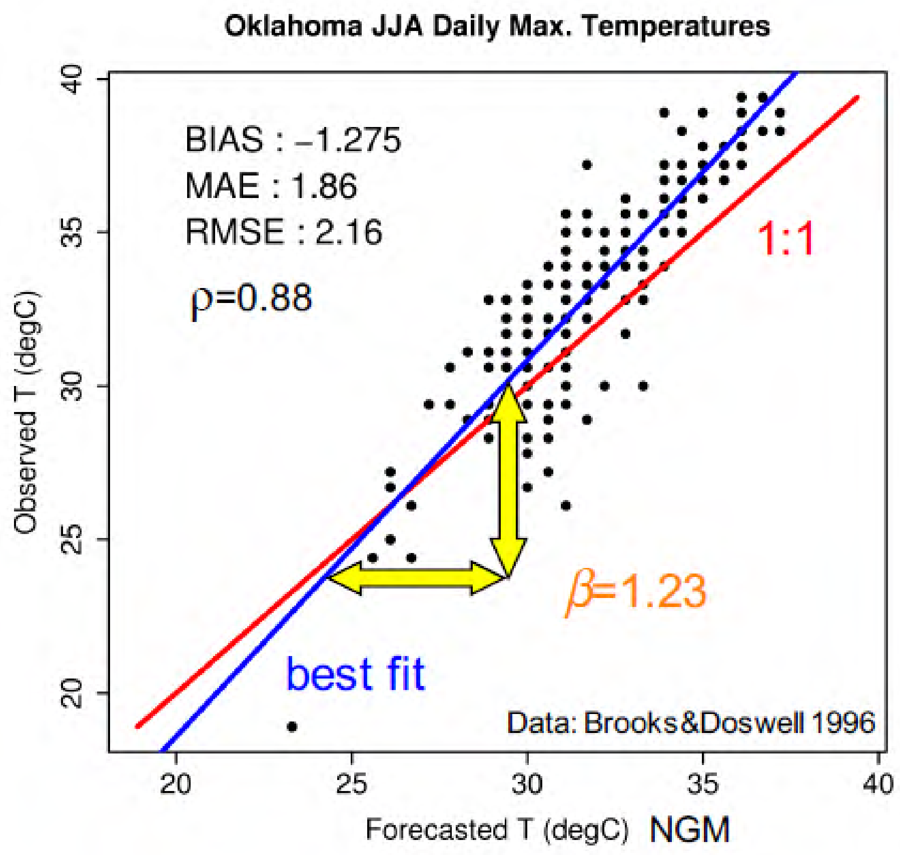
- Linear Regression:

$$o_i = \beta \cdot y_i + \alpha + \varepsilon_i$$

- Linear regression slope

$$\beta = \frac{s_o}{s_y} \cdot \rho$$

- β = 1 best score
- Deviations of β from 1 measure conditional bias
- β is a function of correlation and fraction of variances



MSE Skill Score

$$SS = \frac{MSE - MSE_{clim}}{MSE_{perfect} - MSE_{clim}} = 1 - \frac{MSE}{MSE_{clim}} = 1 - \frac{\frac{1}{N}\sum (y_i - o_i)^2}{s_o^2}$$

- skill score with A=MSE and reference = climatological
- value range: $-\infty < SS \leq 1$;
- perfect forecast: SS = 1; climatology forecast: SS = 0
- random forecast with same variance and mean like observations: SS = –1
- Always: SS ≤ ρ^2

Murphy-Epstein Decomposition

- Decomposition of Skill Score

$$\Rightarrow \quad SS = 1 - \frac{MSE}{MSE_{clim}} = \rho^2 - \underbrace{\left(\rho - \frac{s_y}{s_o}\right)^2}_{\left[\frac{s_y}{s_o}(\beta - 1)\right]^2} - \frac{\left(\overline{y} - \overline{o}\right)^2}{s_o^2}$$

linear correspondence
“maximum explained variance”

penalty for
conditional bias

penalty for
absolute bias

Murphy & Epstein 1989

Accuracy, Association and Skill

- **Accuracy** is a measure of the correspondence between individual pairs of forecasts and observations. MAE and MSE are measures of accuracy.

- **Association** is the overall strength of the relationship between individual pairs of forecasts and observations. The correlation coefficient ρ is thus a measure of linear association.

- **Skill scores** are used to compare the performance of the forecasts with that of a reference forecast such as climatology or persistence.

Probability Forecast

- A probability statement (Probability Density Function (PDF) or Cumulative Distribution Function (CDF)) conveys level of uncertainty of a given forecast
- Categorical forecast: Yes/No. Only 100% or 0% probability
- Probabilistic forecast: assigns a probability value between 0 and 100%
- Example: There is a 30% chance of precipitation for today in Taipei City

Characteristics of Probability Forecasts Verification

- **Reliability**: How well the a priori predicted probability forecast of an event coincides with the a posteriori observed frequency of the event.

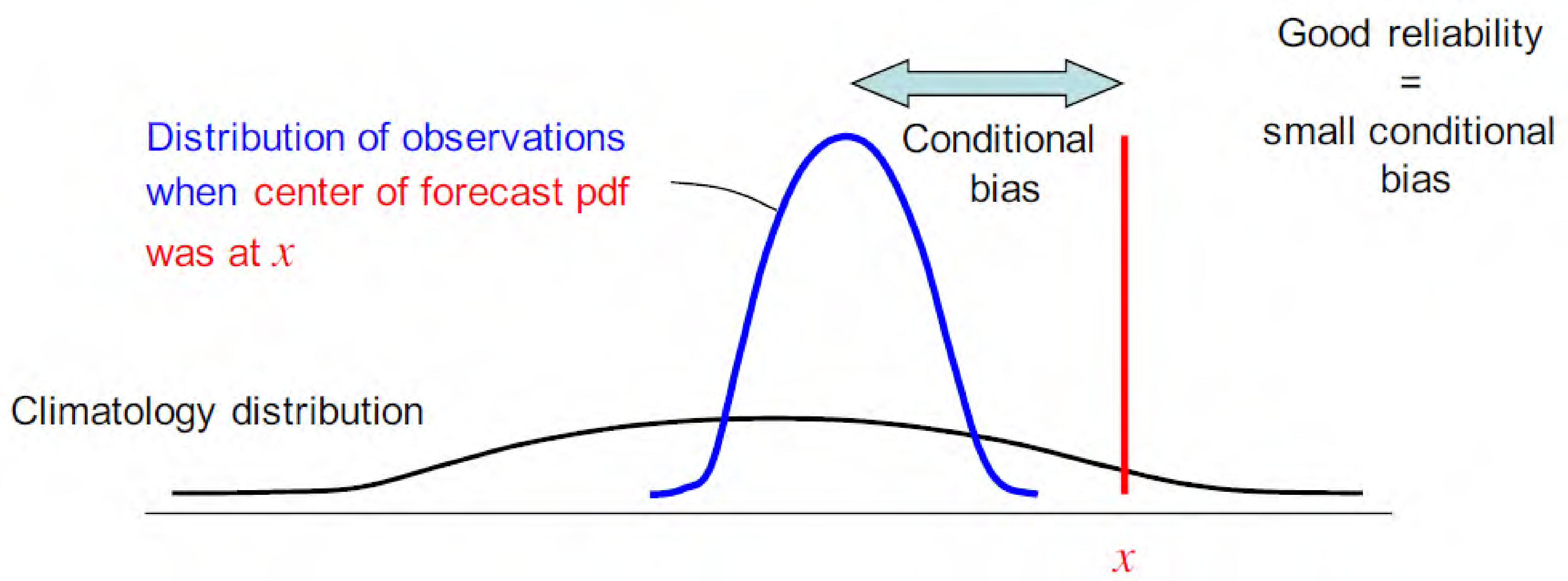
- **Resolution**: How much the forecasts differ from the climatological mean probabilities of the event, and the systems gets it right?

- **Sharpness**: How much do the forecasts differ from the climatological mean probabilities of the event?

- **Skill**: How much better are the forecasts compared to a reference prediction system (chance, climatology, persistence,…)?

Characteristics of Probability Forecasts

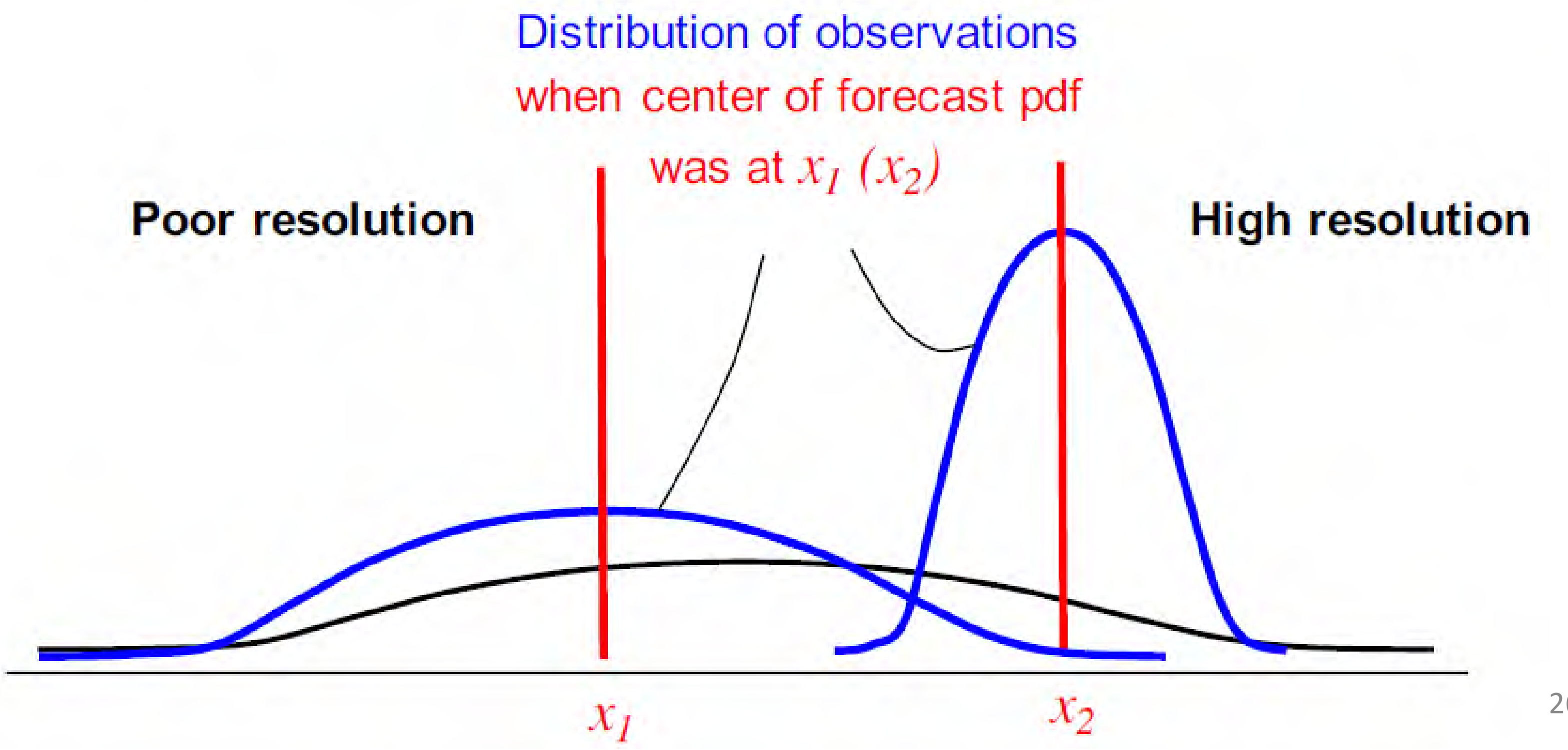
- **Reliability**
  - A measure of systematic and conditional bias
  - High reliability if forecast probability and observed frequency agree



Characteristics of Probability Forecasts

- **Resolution**

  A measure of the ability to distinguish between situations with characteristically different predictands

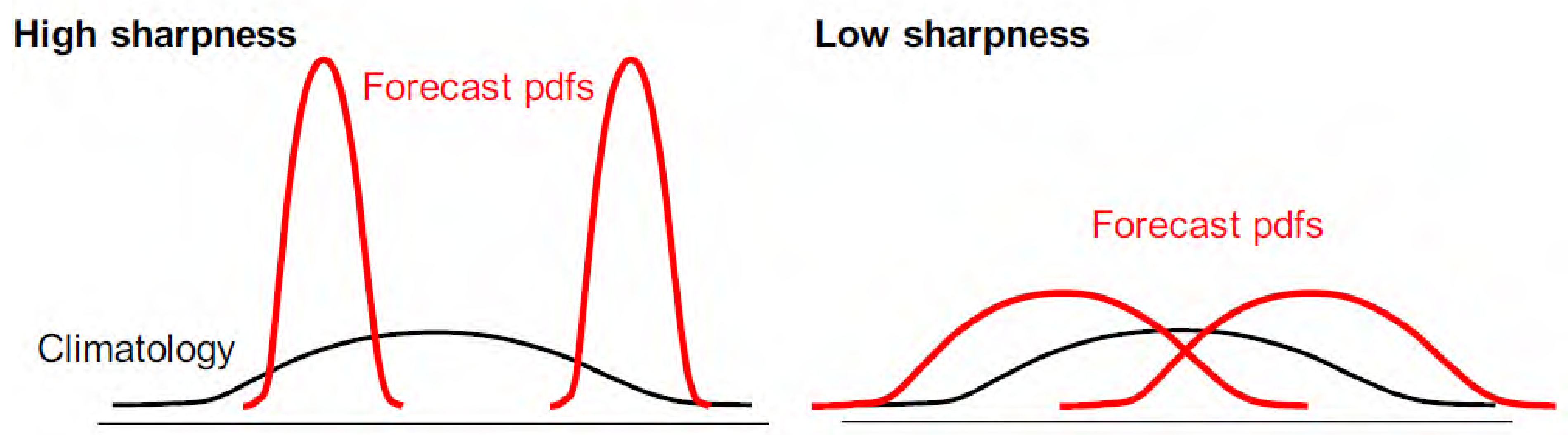


Characteristics of Probability Forecasts

- **Sharpness**

a. A measure of the forecast's confidence

b. Tendency to forecast high probabilities across the whole value range of the predictand



The Brier Score

- **Brier Score**

$$BS = \frac{1}{N}\sum_{i=1}^{N}(Y_i - O_i)^2$$

Y_i: forecasted event probability

O_i: event (1), no-event (0)

$\bar{o}$: climatological event frequency

– Measures accuracy

– Similar to *MSE* but with probabilities

– 0≤ *BS* ≤1; perfect forecast: *BS* = 0

– Climatology forecast: $Y_i = \bar{o} \Rightarrow BS_{clim} = \bar{o}(1-\bar{o})$

Decomposition of Brier Score

$$BS = \sum_{k=1}^{K} \frac{N_k}{N}(y_k - o_k)^2 - \sum_{k=1}^{K} \frac{N_k}{N}(o_k - \overline{o})^2 + \overline{o} \cdot (1 - \overline{o})$$

Reliability:
Weighted average of squared deviations of calibration function from diagonal in reliability plot. Quantifies systematic and conditional biases.

Resolution:
Weighted average of squared deviations of calibration function from climatology in refinement diagram. Ability to discern between events / no-events.

Uncertainty:
Depends on climatology only (not on forecast). Binomial variance. Measures the uncertainty of the forecasting situation. Maximum at $\overline{o}$=0.5.

Brier Skill Score

$$BSS = \frac{BS - BS_{ref}}{BS_{perf} - BS_{ref}} = 1 - \frac{BS}{BS_{ref}}$$

– Generic form of skill scores.

– Perfect: *BSS* = 1, no-skill: *BSS* ≤ 0

– E.g. $BS_{ref} = BS_{clim}$

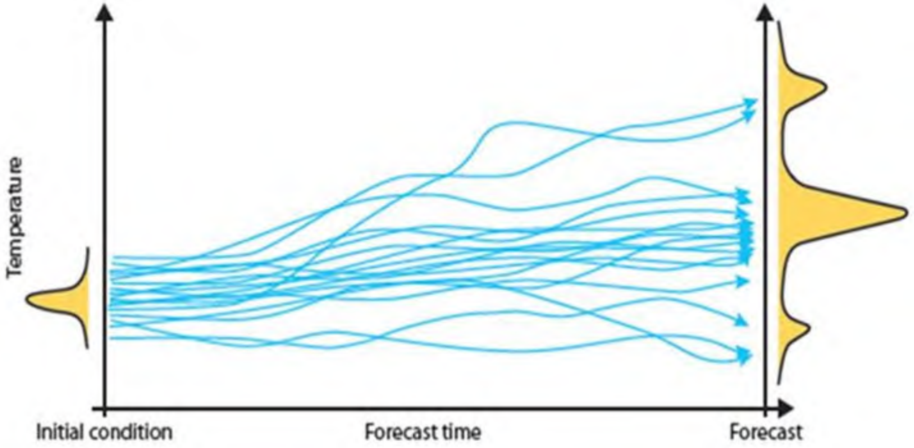
- **Decomposition of Brier Skill Score**

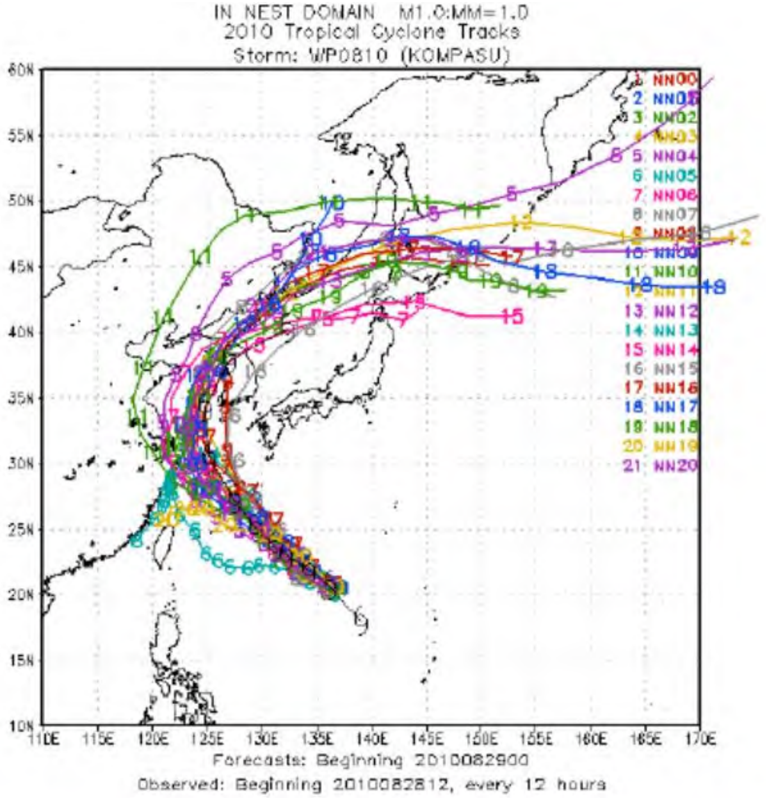
$$BSS = 1 - \frac{BS}{BS_{clim}} = \frac{Resolution - Reliability}{Uncertainty}$$

Ensemble Forecast

- Ensemble forecasting methods involve evaluating a set of runs from an Numerical Weather Prediction (NWP) model, or different NWP models, from the same initial time.

- Ensemble forecasting is a form of Monte Carlo analysis.

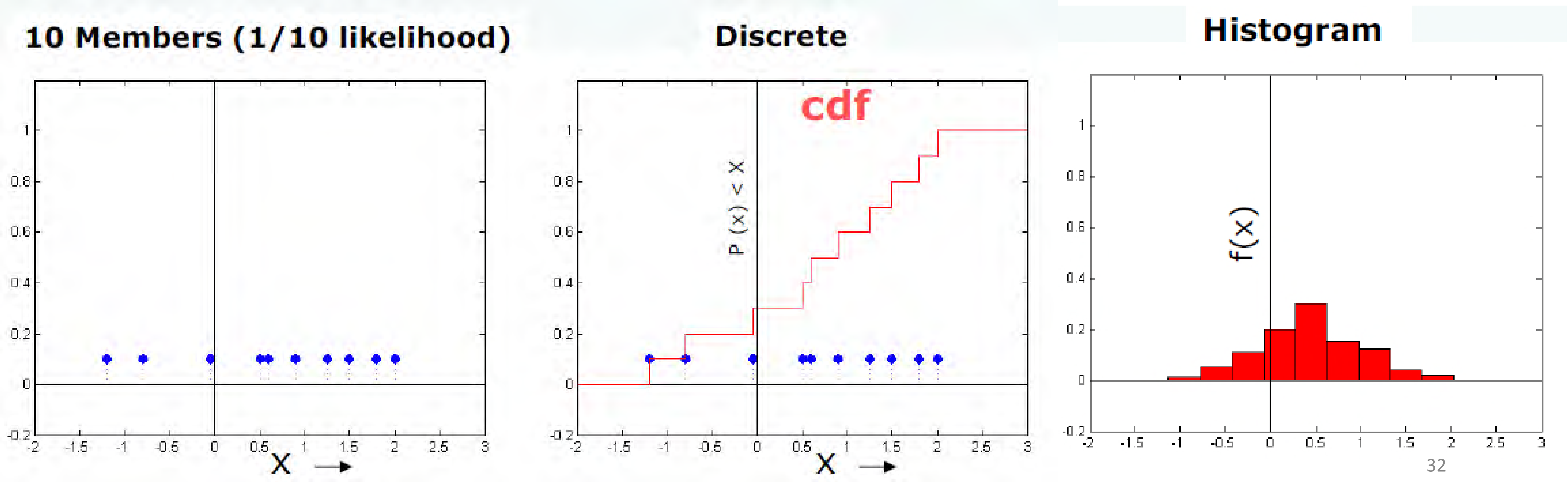




Hurricane tracks

Ensemble Distribution

- Ensemble forecast will form a probability distribution reflecting the uncertainty associated with initial and model errors.

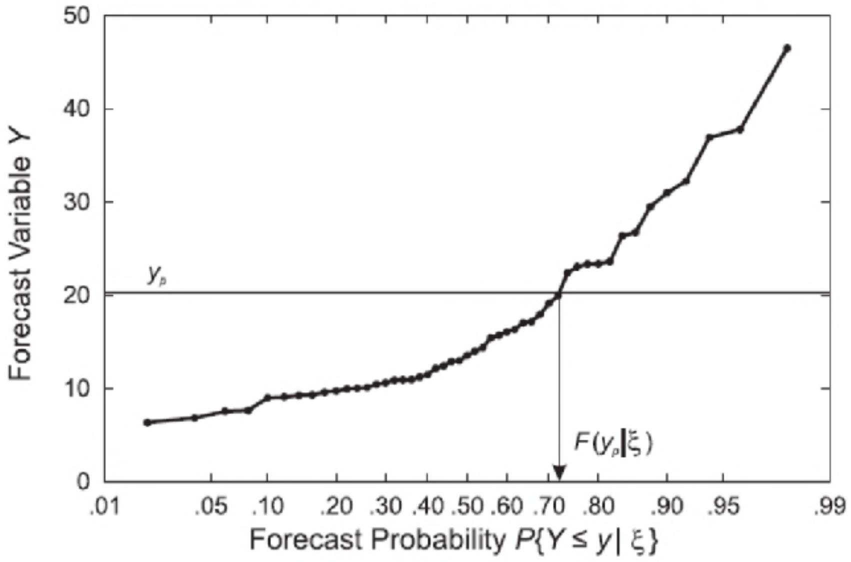


Verification of Ensemble Forecasts

- **Mean Square Error** (**Brier Score**)

$$MSE(y_p) = E\{F(\mathrm{y}_p|\xi) - X(y_p)\}^2$$

– y_p: the threshold
– $F(\mathrm{y}_p|\xi) = P\{Y \leq \mathrm{y}_p|\xi\}$; ensemble forecast probability
– $X(y_p) = 1$ if $\{\mathrm{Y} \leq \mathrm{y}_p\}$ and $X(y_p) = 0$ if otherwise



- **Estimated the forecast probability**

$$f_t(y_{p_i}) = \frac{1}{M_t}\sum_{j=1}^{M_t} I[y_{p_i} - z_t(j)]$$

– $z_t(j),\ j = 1, \cdots, M_t$: the ensemble forecast at time t

Brier Skill Score (SS)

$$\mathrm{SS}(\mathrm{y}_p) = 1 - \frac{MSE(y_p)}{\sigma_x^2(y_p)}$$

where $\sigma_x^2(\mathrm{y}_p) = p(1-p)$

- **Skill Score Decomposition**

$$\mathrm{SS} = \rho_{fx}^2 - \left[\rho_{fx} - \left(\frac{\sigma_f}{\sigma_x}\right)\right]^2 - \left[\frac{(\mu_f - \mu_x)}{\sigma_x}\right]^2$$

$$\mathrm{SS}(p) = \mathrm{PS}(p) - \mathrm{CB}(p) - \mathrm{UB}(p)$$

- SS(*P*): Skill function
- PS(*p*): Potential skill function
- CB(*p*): Conditional bias function
- UB(*p*): Unconditional bias function

Average Forecast Quality for Discrete Variable

- **Ranked Probability Score(RPS)**

  Measures the quadratic distance between forecast and verification probabilities for several probability categories k.

$$\overline{\mathrm{RPS}} = \frac{1}{k}\sum_{i=1}^{k} MSE(y_i)$$

MSE(*y*): Brier Score in discrete forecast *y*

- **Ranked Probability Skill Score (RPSS)**

  A measure for skill relative to a reference forecast

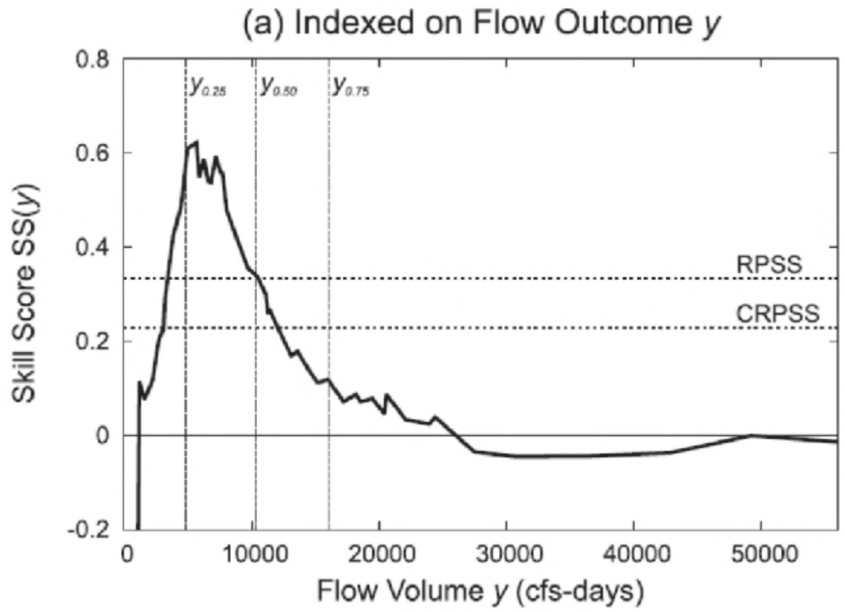
$$\mathrm{RPSS} = 1 - \frac{\overline{RPS}}{\overline{RPS_c}} = 1 - \frac{\sum_{i=1}^{k} MSE(y_i)}{\sum_{i=1}^{k} \sigma_x^2 (y_i)} = 1 - \frac{\sum_{i=1}^{k} \sigma_x^2 (y_i) SS(y_i)}{\sum_{i=1}^{k} \sigma_x^2 (y_i)}$$

Average Forecast Quality for Continuous Variable

- **Continuous Ranked Probability Score (CRPS)**

$$\overline{\text{CRPS}} = \int_{-\infty}^{\infty} MSE(y)dy$$

MSE(*y*):Brier Score in continuous forecast *y*



- **Continuous Ranked Probability Skill Score (CRPSS)**

$$\text{CRPSS} = 1 - \frac{\overline{CRPS}}{\overline{CPRS_c}} = \int_{-\infty}^{\infty} w(y)SS(y)dy \quad ,\text{w(y)} = \frac{\sigma_x^2(y)}{\int_{-\infty}^{\infty} \sigma_x^2(y)dy}$$

Summary Measures using Probability Thresholds

- Average **Brier Score** using probability *p*

$$\overline{\text{MSE}} = \int_0^1 MSE(p)dp$$

- Average **Brier Skill Score** in probability *p*

$$\overline{\text{SS}}_p = \int_0^1 w(p)SS(p)dp$$

$$w(p) = \frac{p(1-p)}{\int_0^1 p(1-p)dp}$$, the weight function

Summary Measure_ Mass $\bar{Q}_p$

- **The weighted-average forecast quality function**

$$\bar{Q}_p = \int_0^1 w(p)Q(p)dp = \int_0^1 M(p)dp$$

M(*p*) is a mass distribution

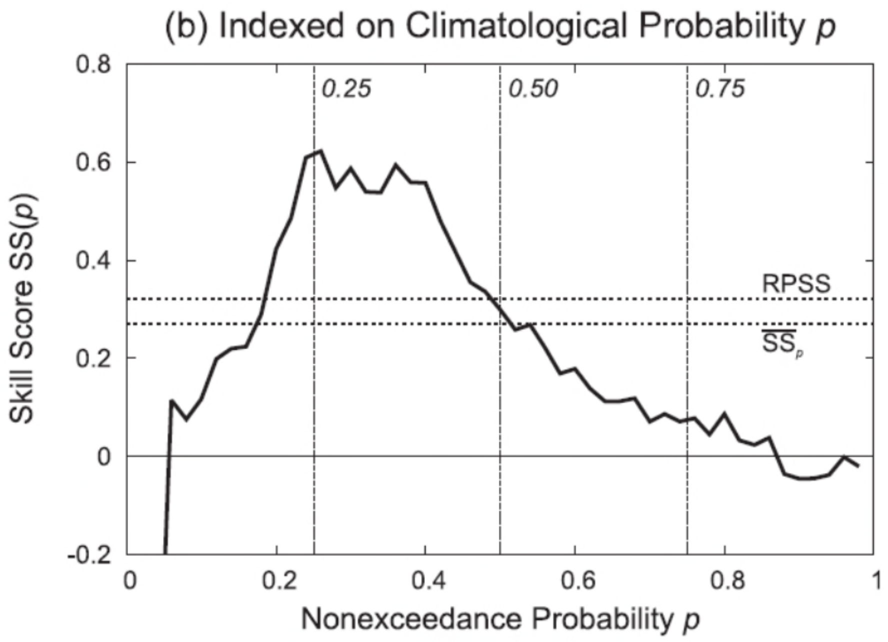
– *Q*(*p*)= SS(*p*), PS(*p*), CB(*p*), UB(*p*)

- **The approximation of $\overline{\boldsymbol{Q}_p}$**

$$\bar{Q}_p = \sum_{i=1}^{k} w(p_i)Q(p_i)$$

– $w(p_i) = \frac{p_i(1-p_i)}{\sum_{i=1}^{k} p_i(1-p_i)}$, the weight function

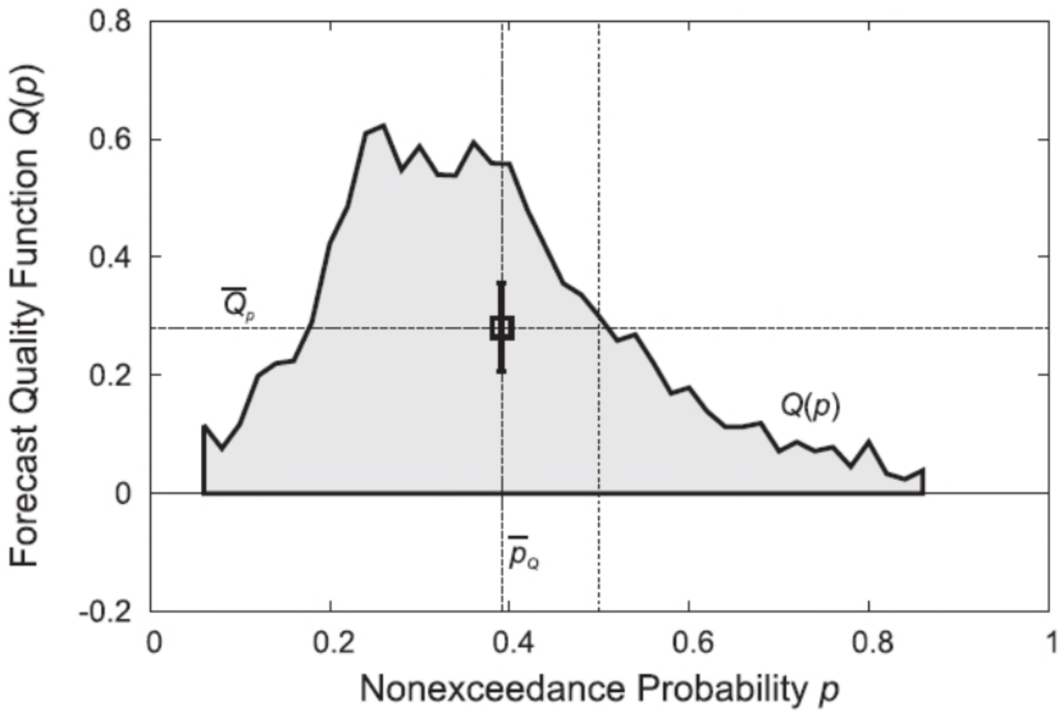
– $\{p_i, i = 1, \cdots, k\}$ be the probability thresholds



Summary Measure_ Center of Mass $\bar{p}_Q$

The location of the center of mass for the weighted forecast quality function

$$\bar{p}_Q = \frac{\int_0^1 pM(p)dp}{\int_0^1 M(p)dp} = \frac{1}{\bar{Q}_p}\int_0^1 pw(p)Q(p)dp$$



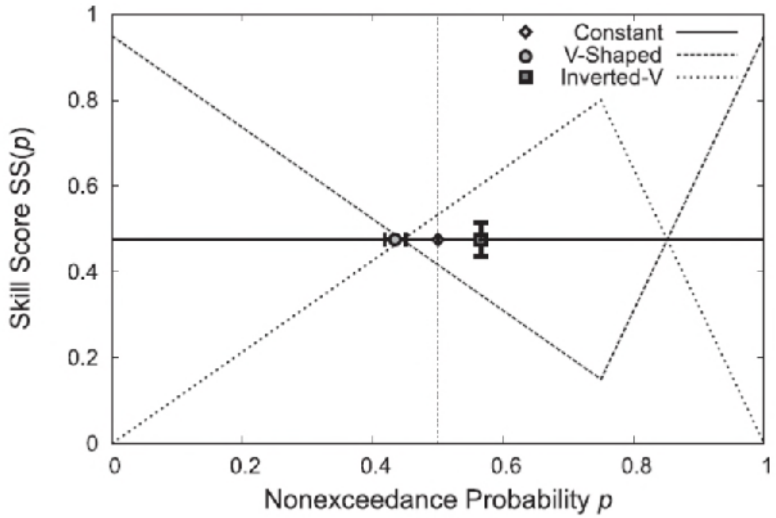
- **The approximation of $\bar{p}_Q$**

$$\bar{p}_Q = \frac{1}{\bar{Q}_p}\sum_{i=1}^{k} p_i w(p_i) Q(p_i)$$

Summary Measure_ Shape Measure γ_{Q_p}

- The Measure of the distribution of mass *M*(*p*) (**shape measure**)

$$\gamma_{Q_p} = k_{Q_p} - \frac{1}{\sqrt{20}}$$



Where $k_{Q_p} = \sqrt{\frac{I_{Q_p}}{\bar{Q}_p}}$, the radius of gyration;

I_{Q_p} is the moment of inertia

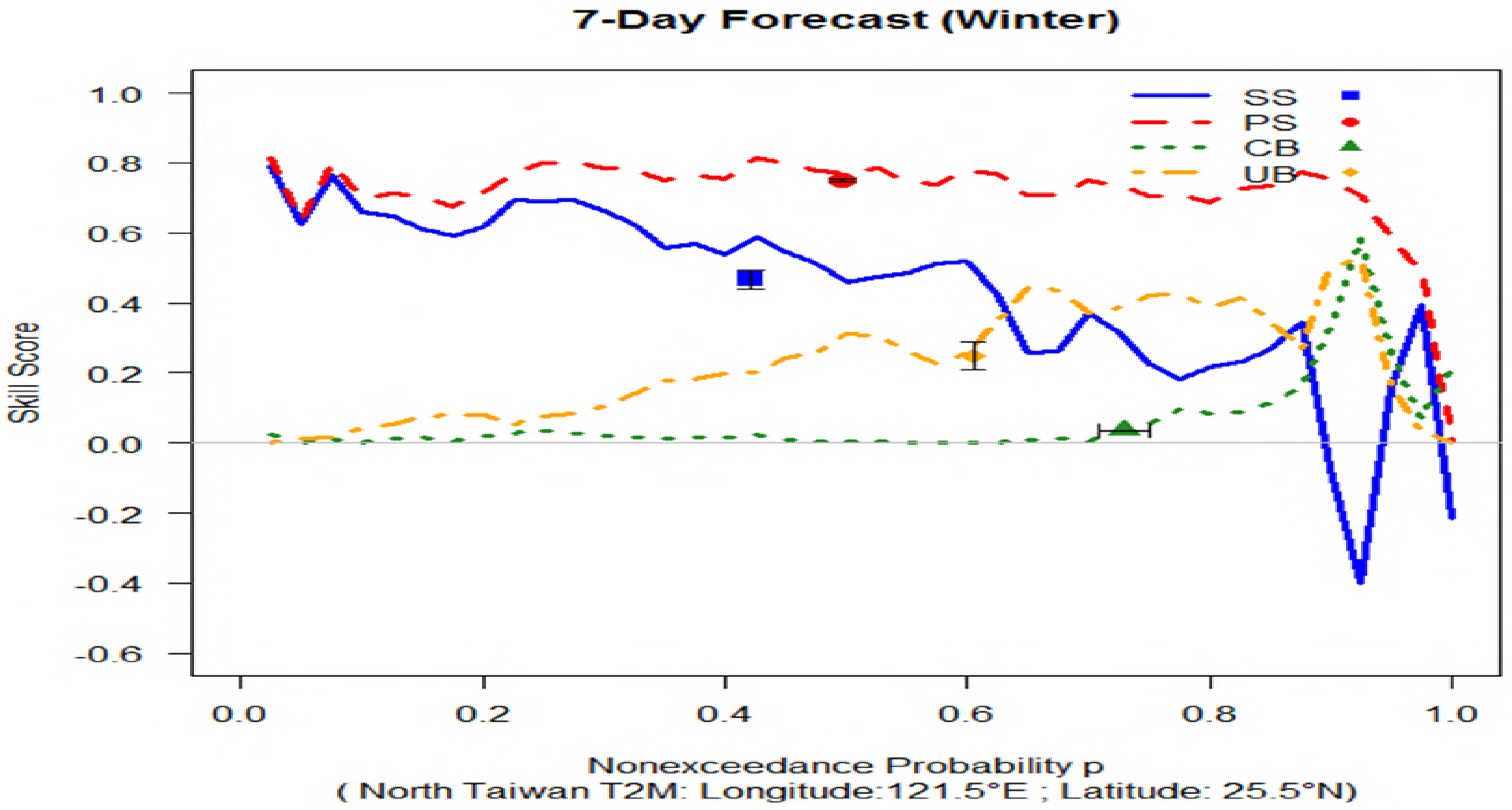
$I_{Q_p} = \int_0^1 p^2 w(p) Q(p) dp - \bar{p}_Q^2\, \bar{Q}_p$

- $\boldsymbol{\gamma_{Q_p} < 0}$ (inverted-V shaped): skill is higher near the center of mass
- $\boldsymbol{\gamma_{Q_p} > 0}$ (V shaped): skill is higher in the extremes

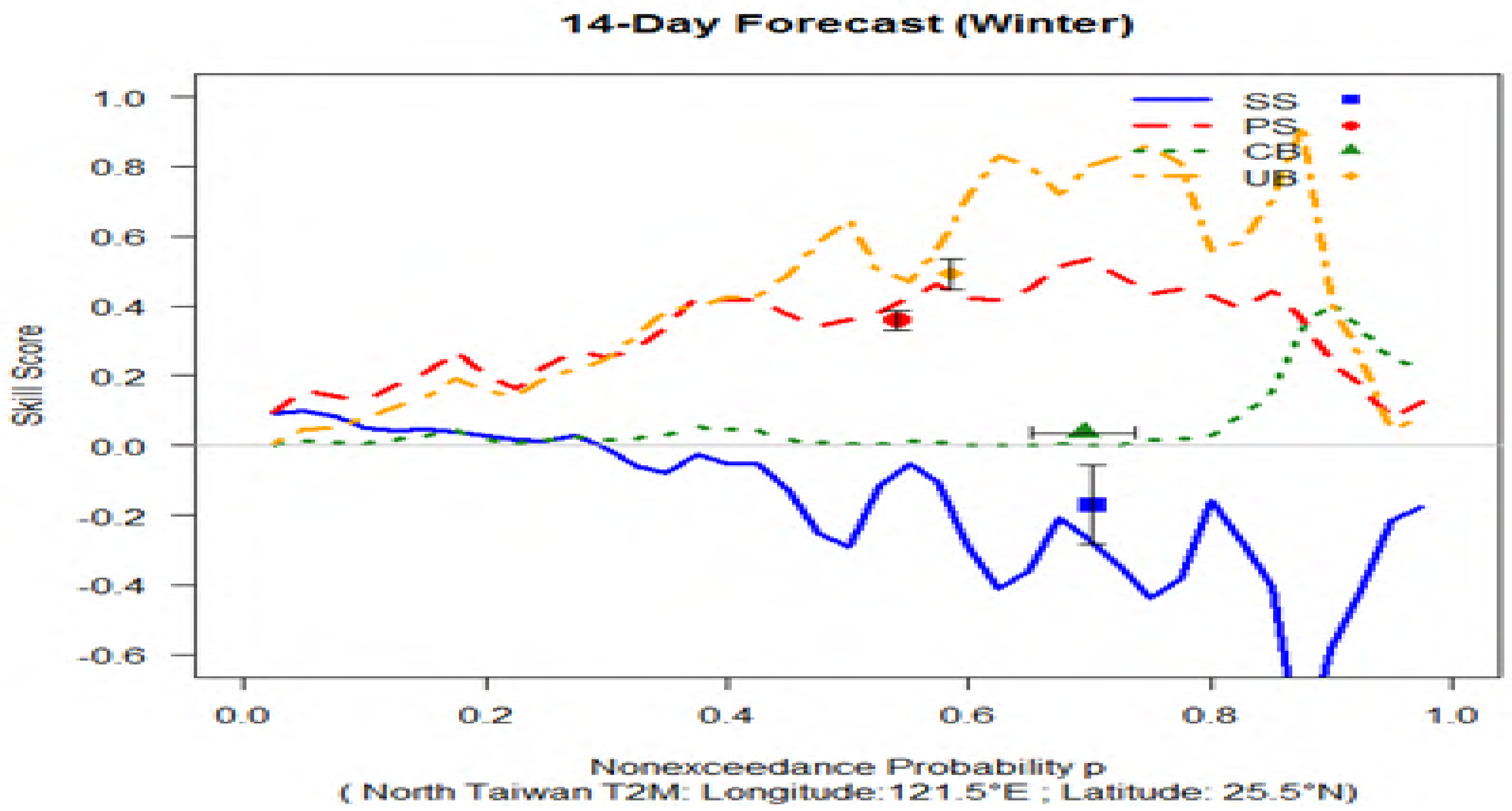
- The approximation of I_{Q_p}

$$I_{Q_p} = \sum_{i=1}^{k} p_i^2 w(p_i) Q(p_i) - \bar{p}_Q^2\, \bar{Q}_p$$

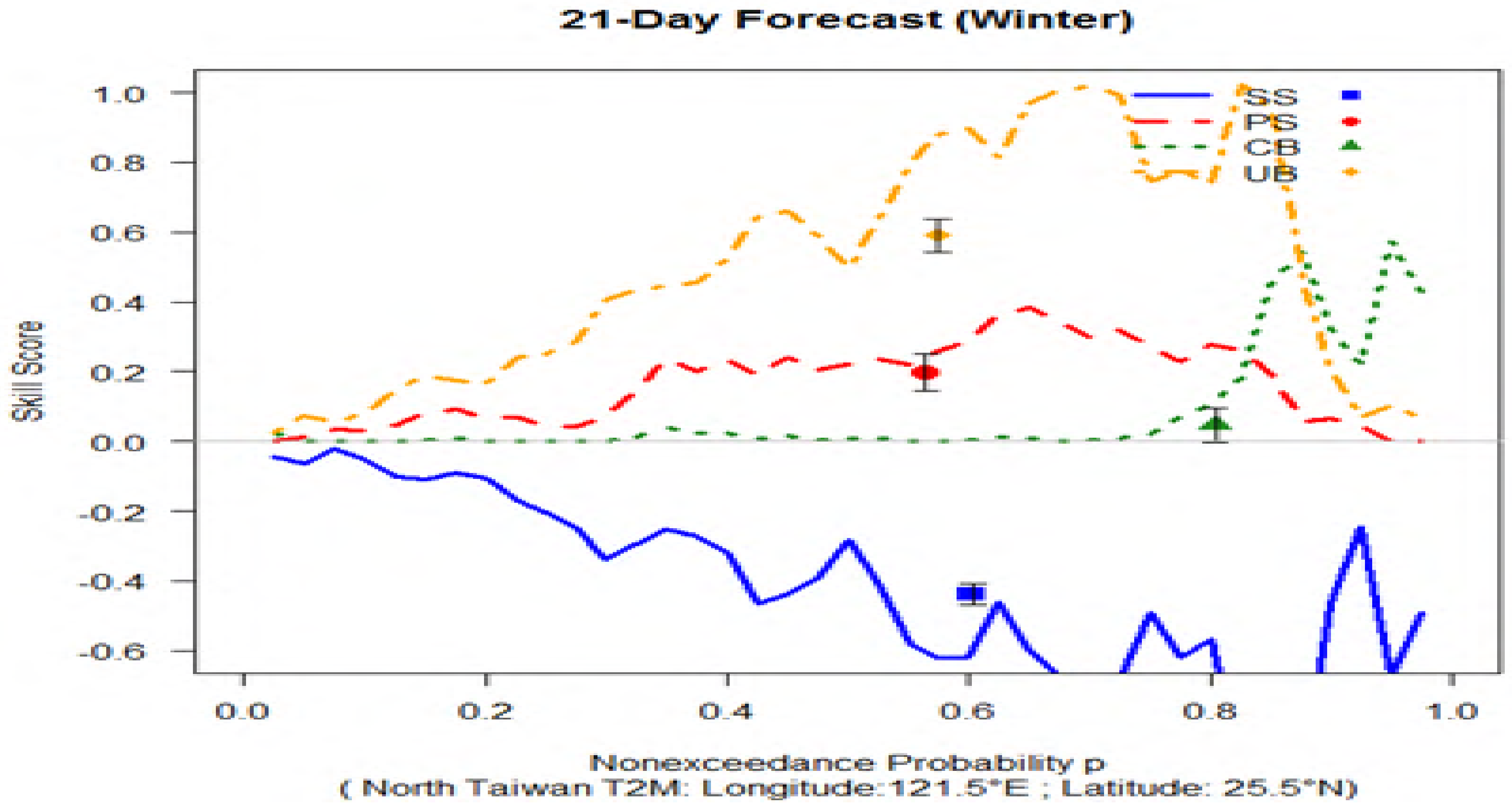
Skill Function and Its Decomposition



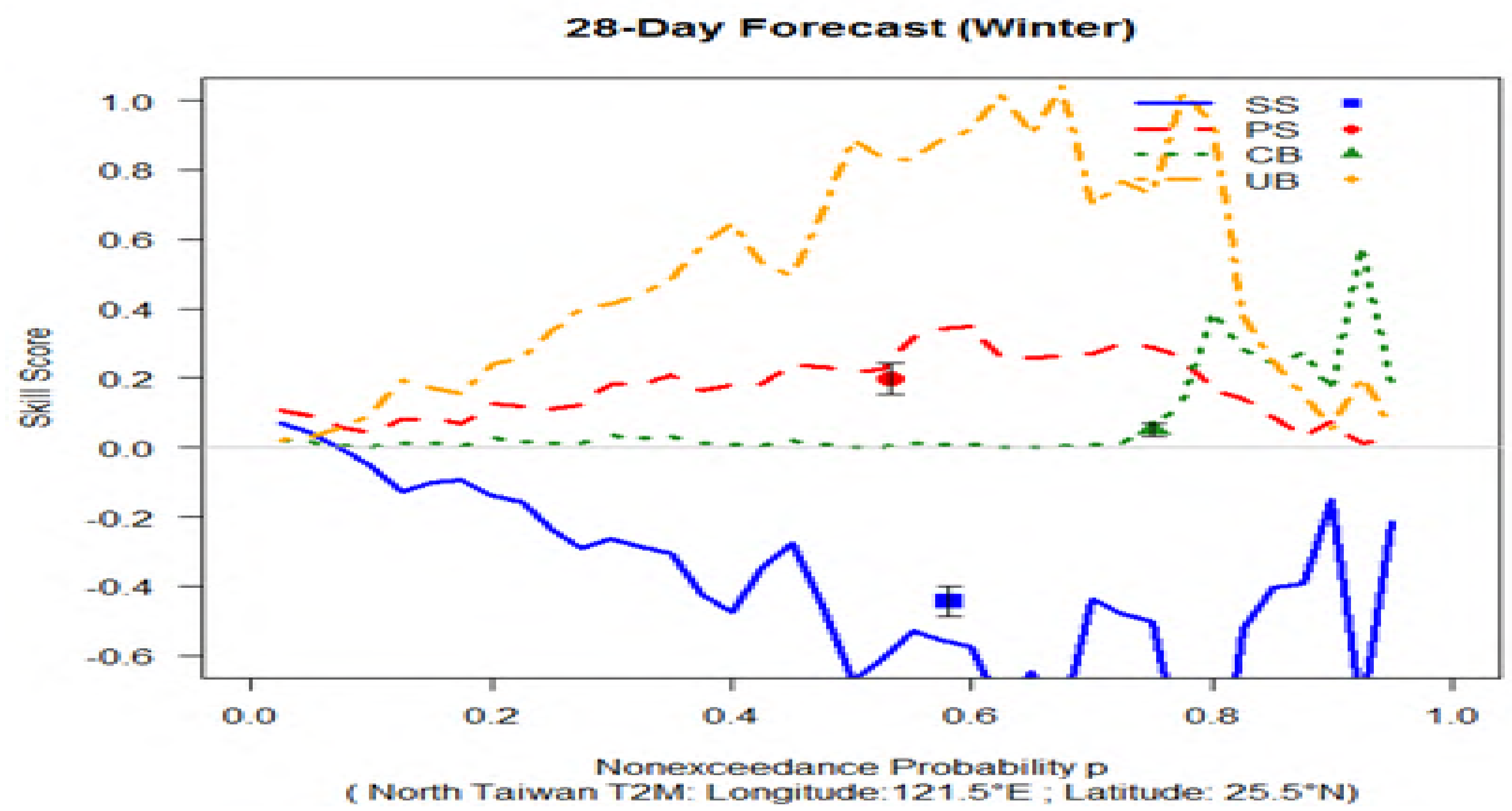
Skill Function and Its Decomposition



Skill Function and Its Decomposition



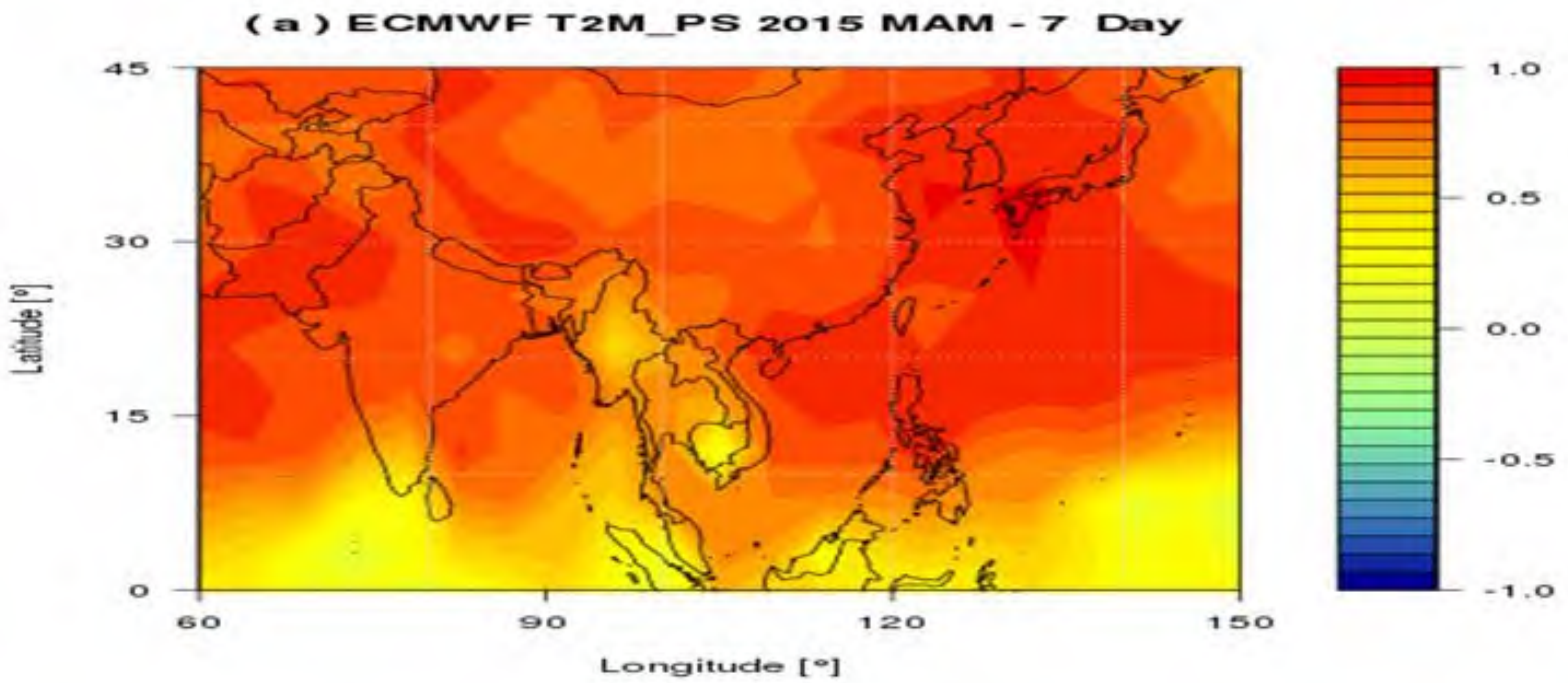
Skill Function and Its Decomposition



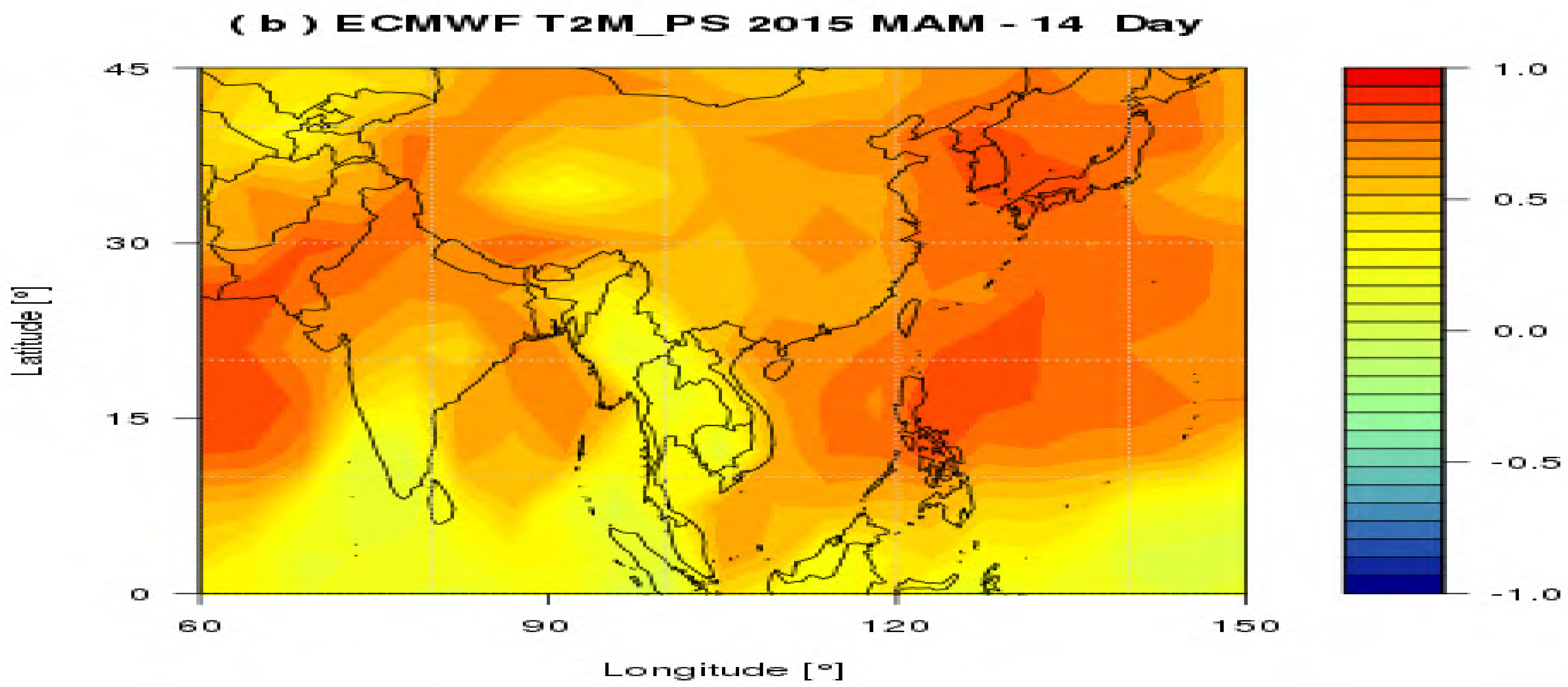
ECMWF S2S Data

- Variables: **T2M, MSLP, TOTPRCP**
- Lead Times: 7 Days, 14 Days, 21 Days, 28 Days
- Study Area: (latitude: 0°N-45°N ; longitude:60°E-150°E)
- Study Seasons: 2015 Spring, 2015 Summer, 2015 Fall, 2015-2016 Winter, 2016 Spring

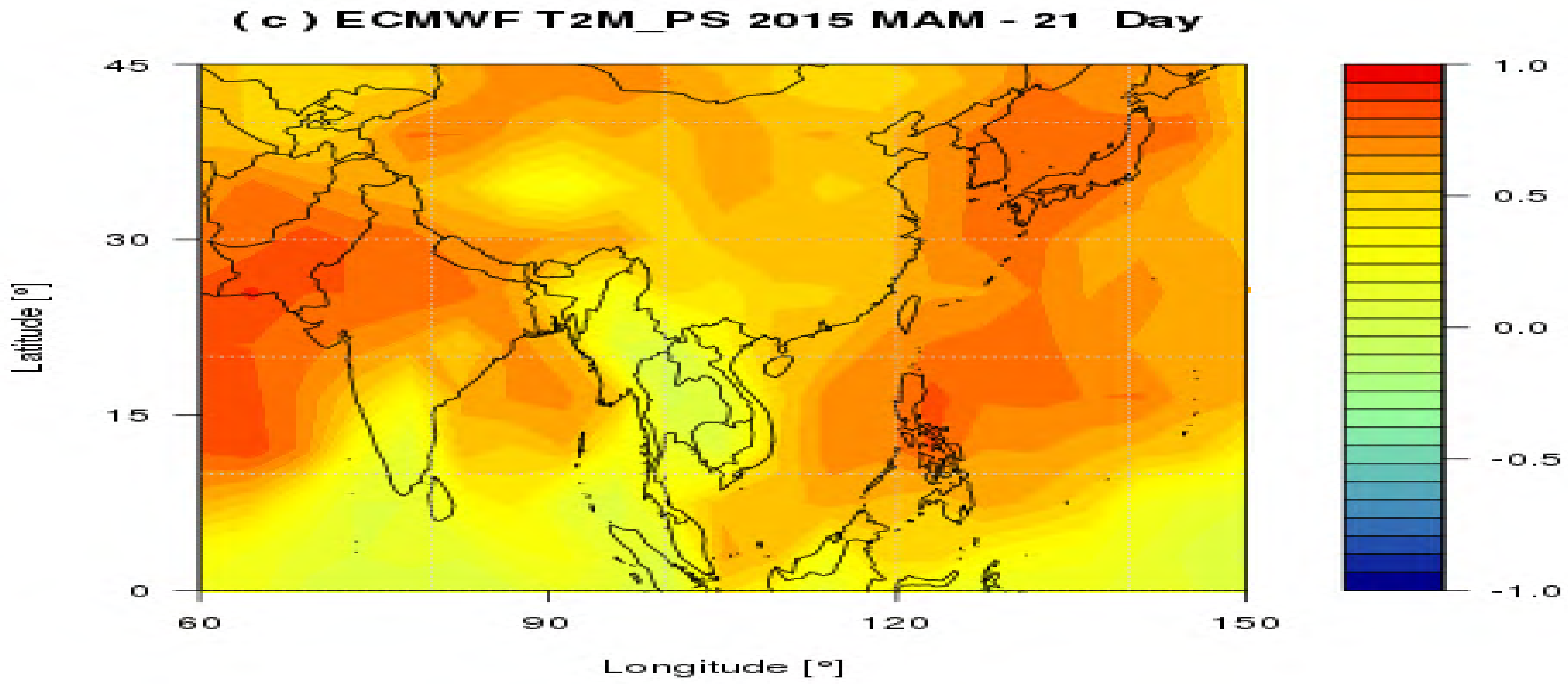
ECMWF S2S 2015 Spring T2M PS



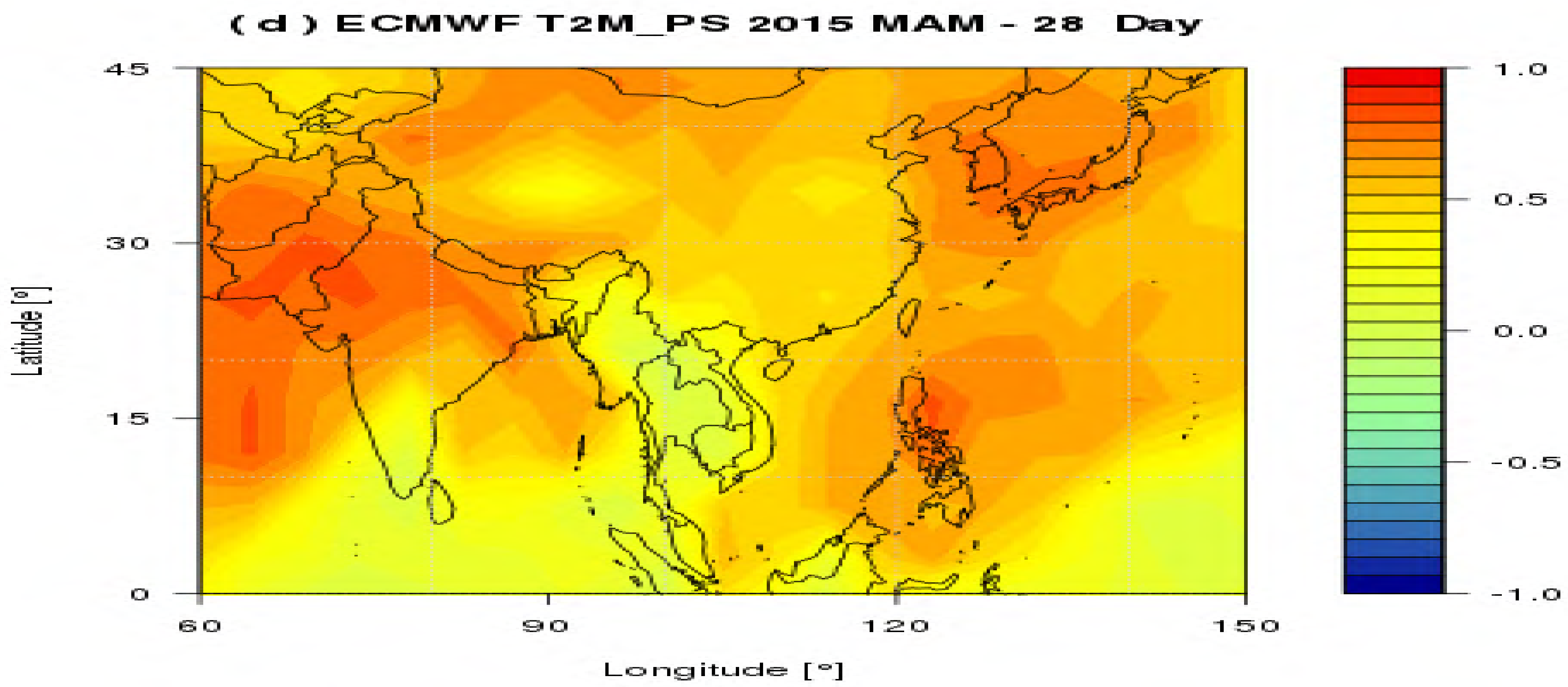
ECMWF S2S 2015 Spring T2M PS



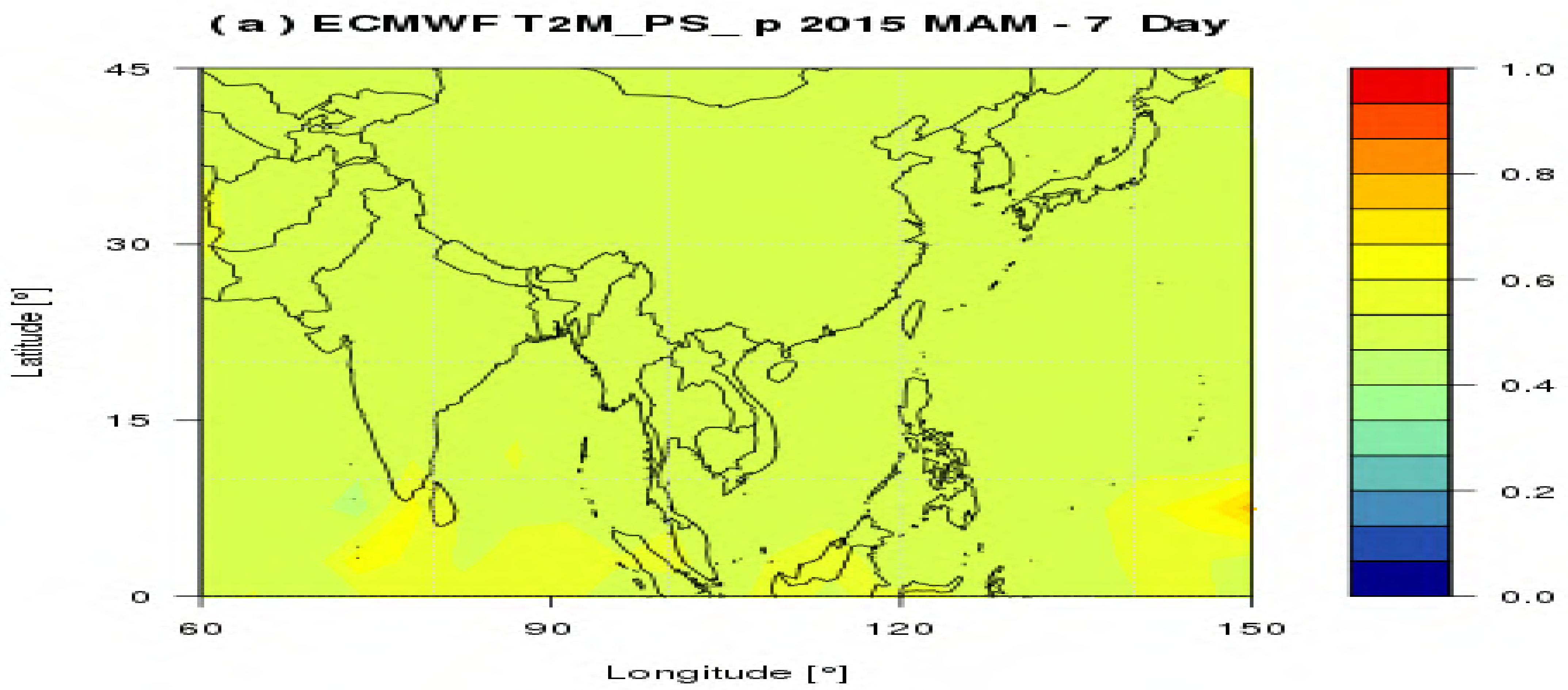
ECMWF S2S 2015 Spring T2M PS



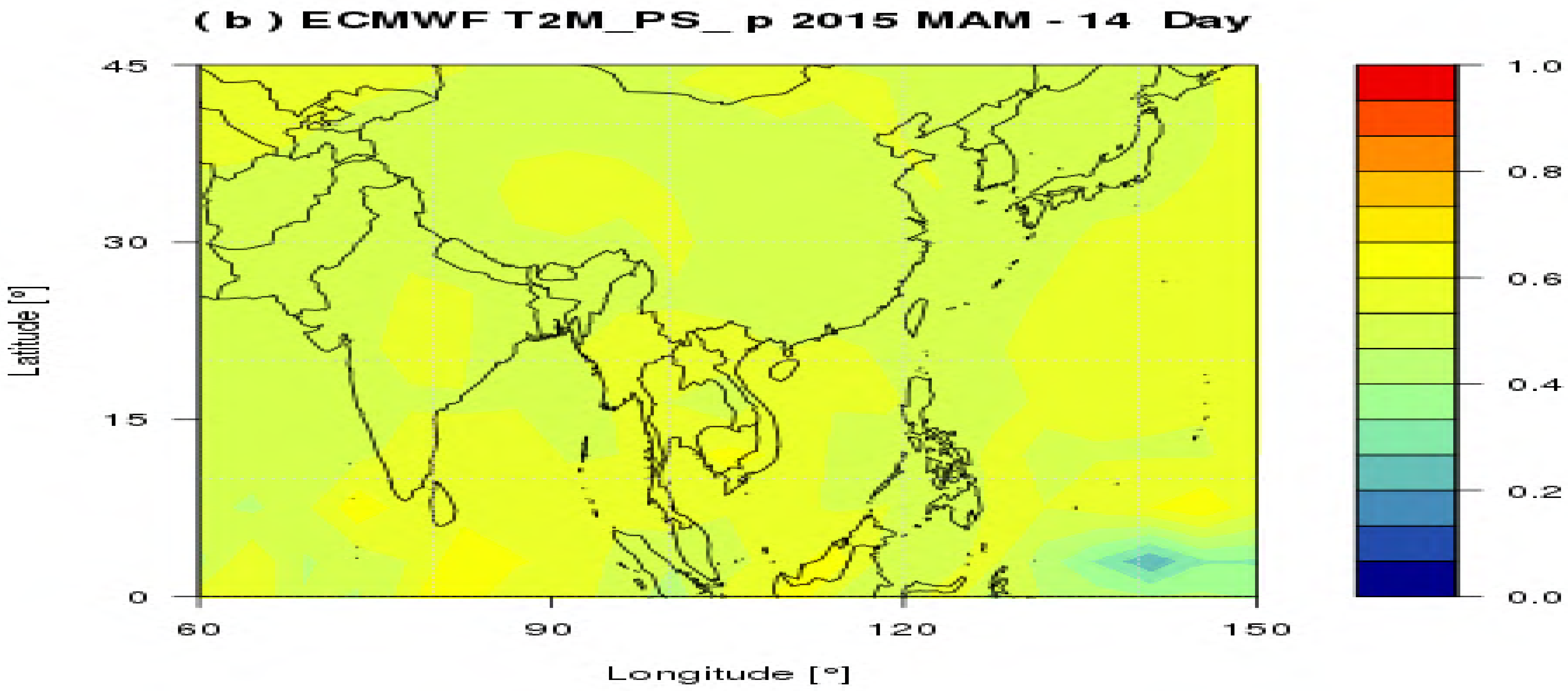
ECMWF S2S 2015 Spring T2M PS



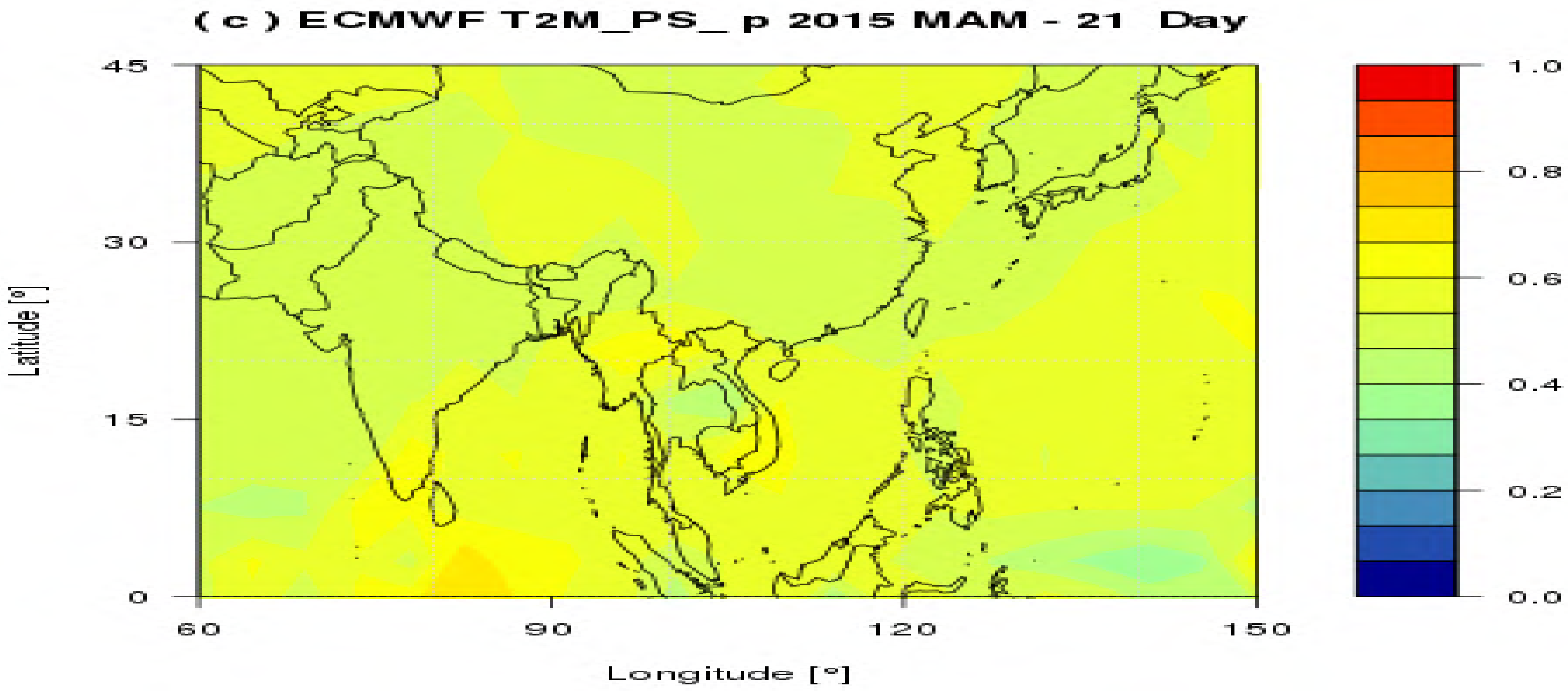
ECMWF S2S 2015 Spring T2M PS_ $\bar{p}_Q$



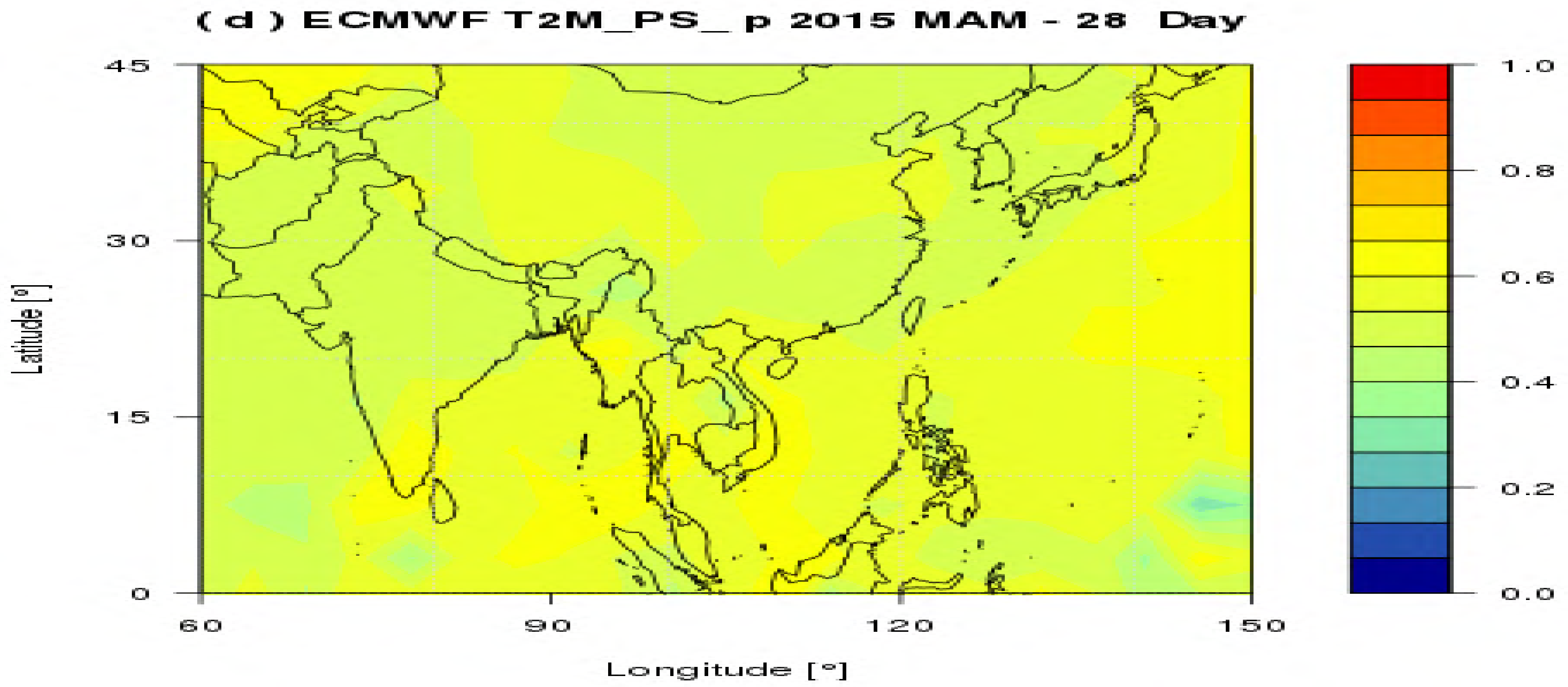
ECMWF S2S 2015 Spring T2M PS_ $\bar{p}_Q$



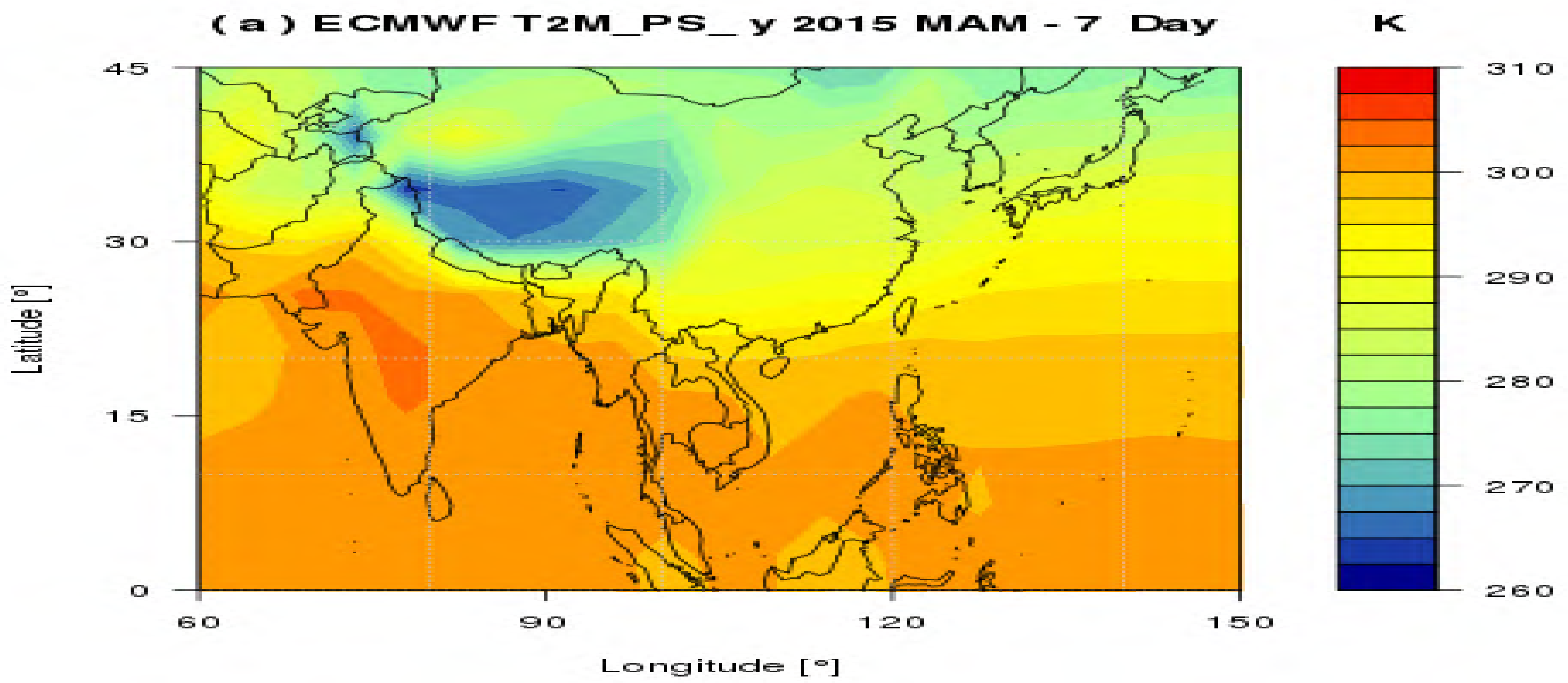
ECMWF S2S 2015 Spring T2M PS_ $\bar{p}_Q$



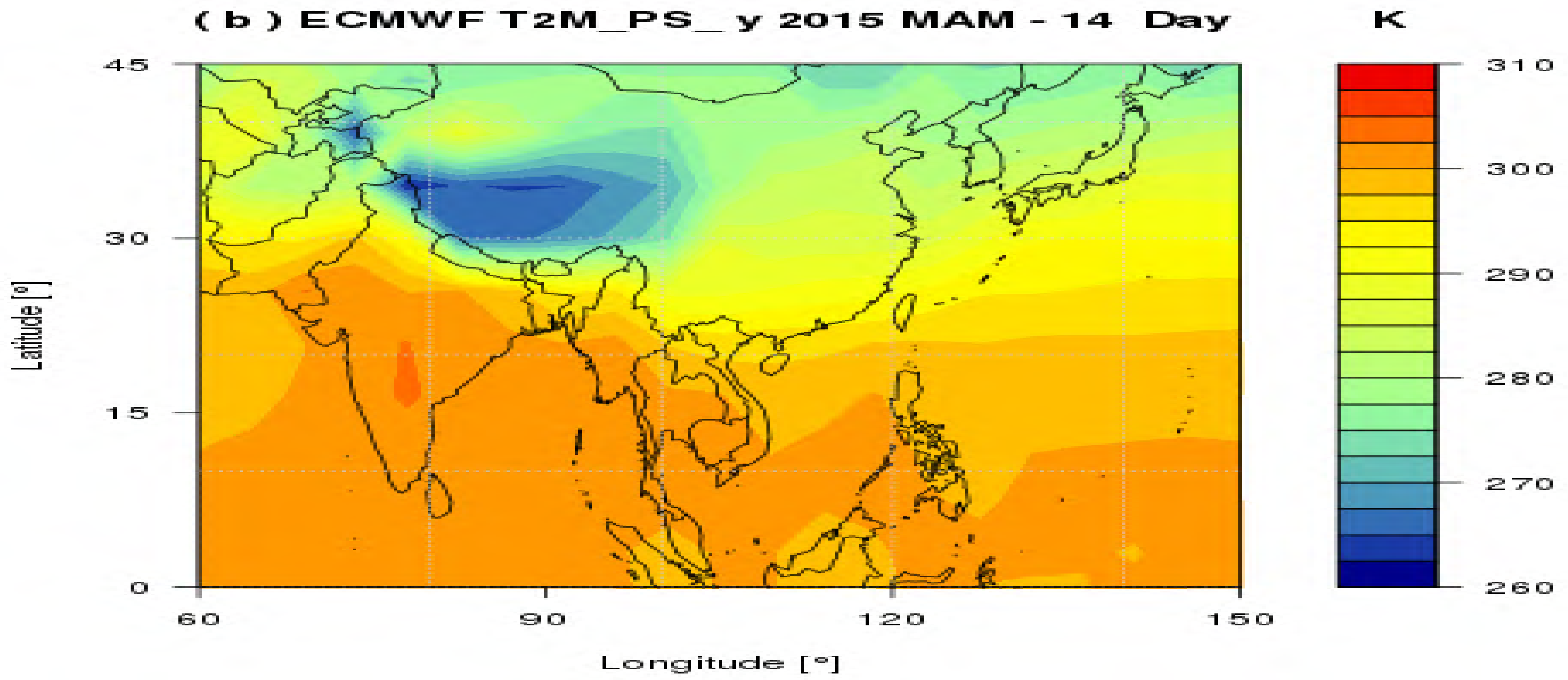
ECMWF S2S 2015 Spring T2M PS_ $\bar{p}_Q$



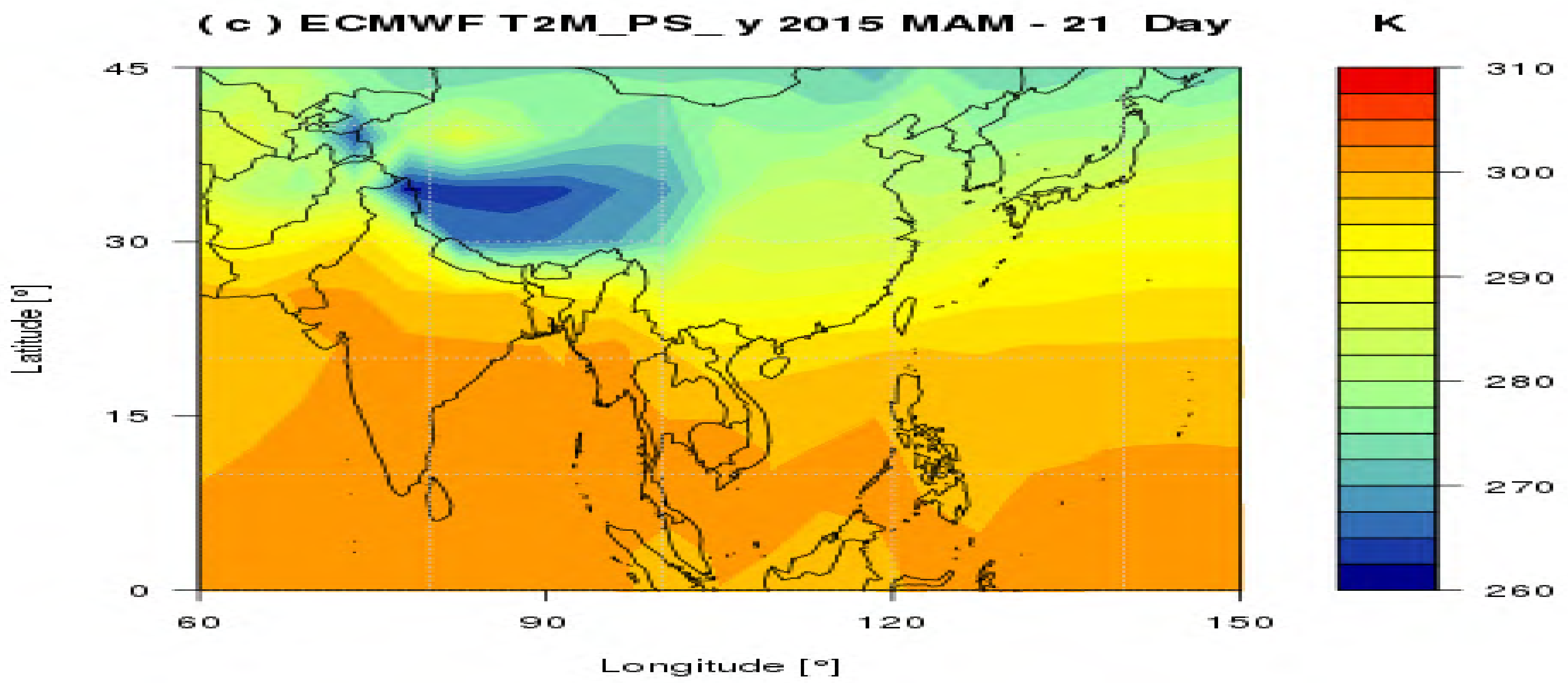
ECMWF S2S 2015 Spring T2M PS_ $\bar{p}_Q(y)$



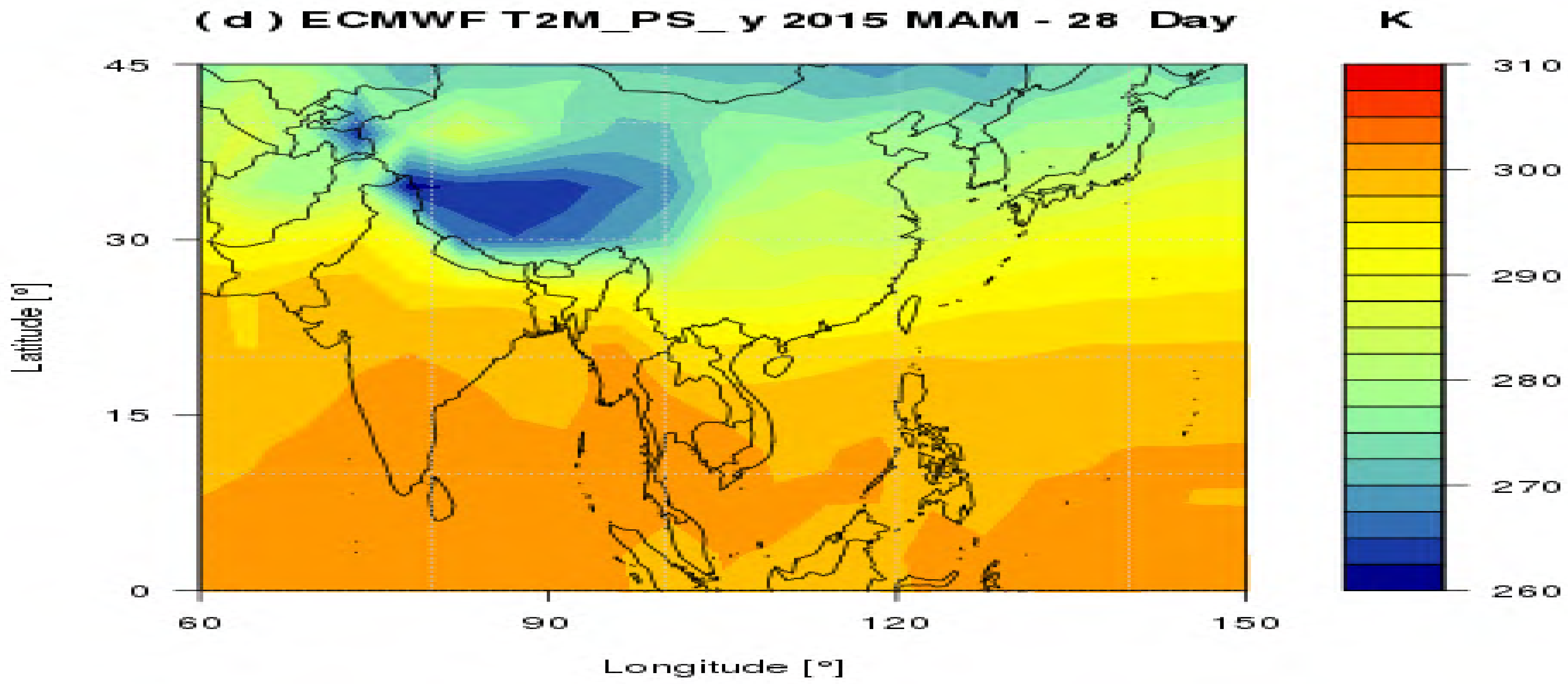
ECMWF S2S 2015 Spring T2M PS_ $\bar{p}_Q$(y)



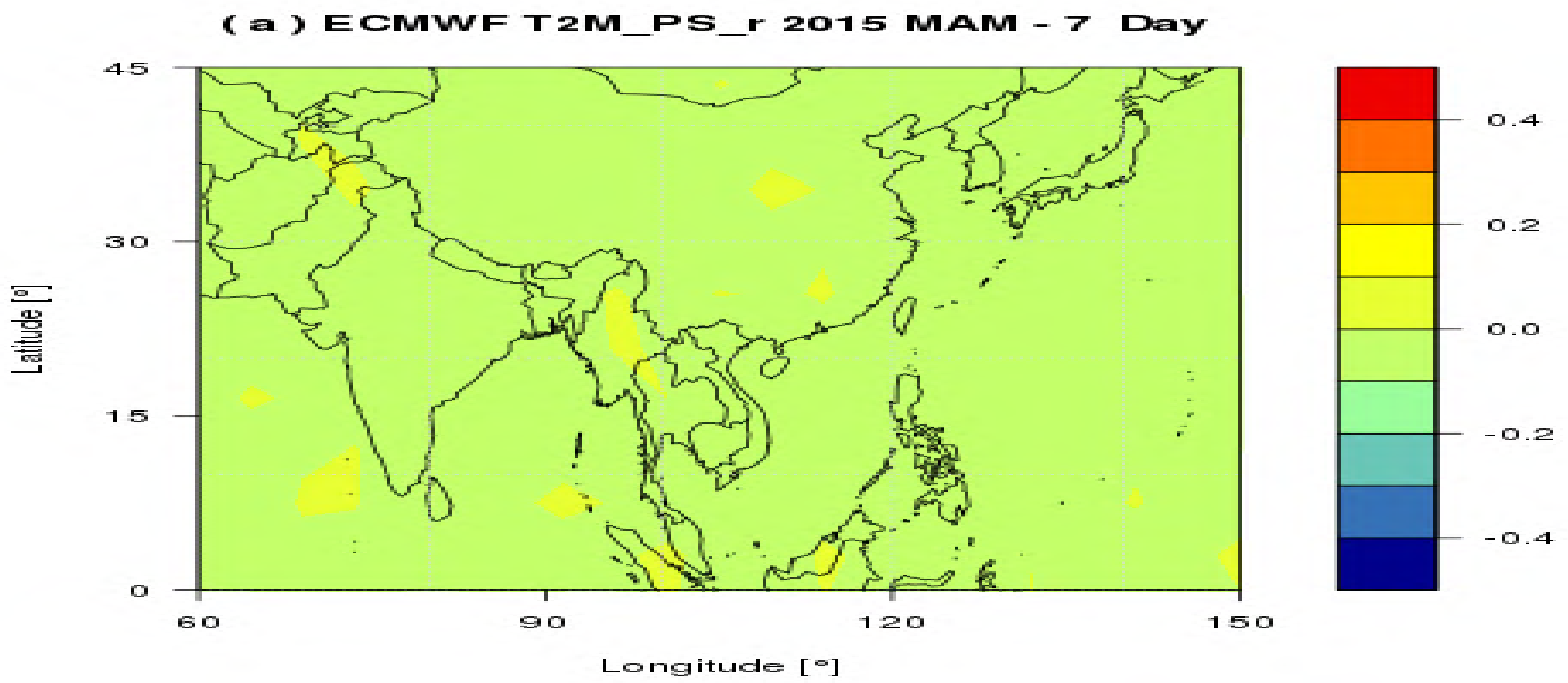
ECMWF S2S 2015 Spring T2M PS_ $\bar{p}_Q$(y)



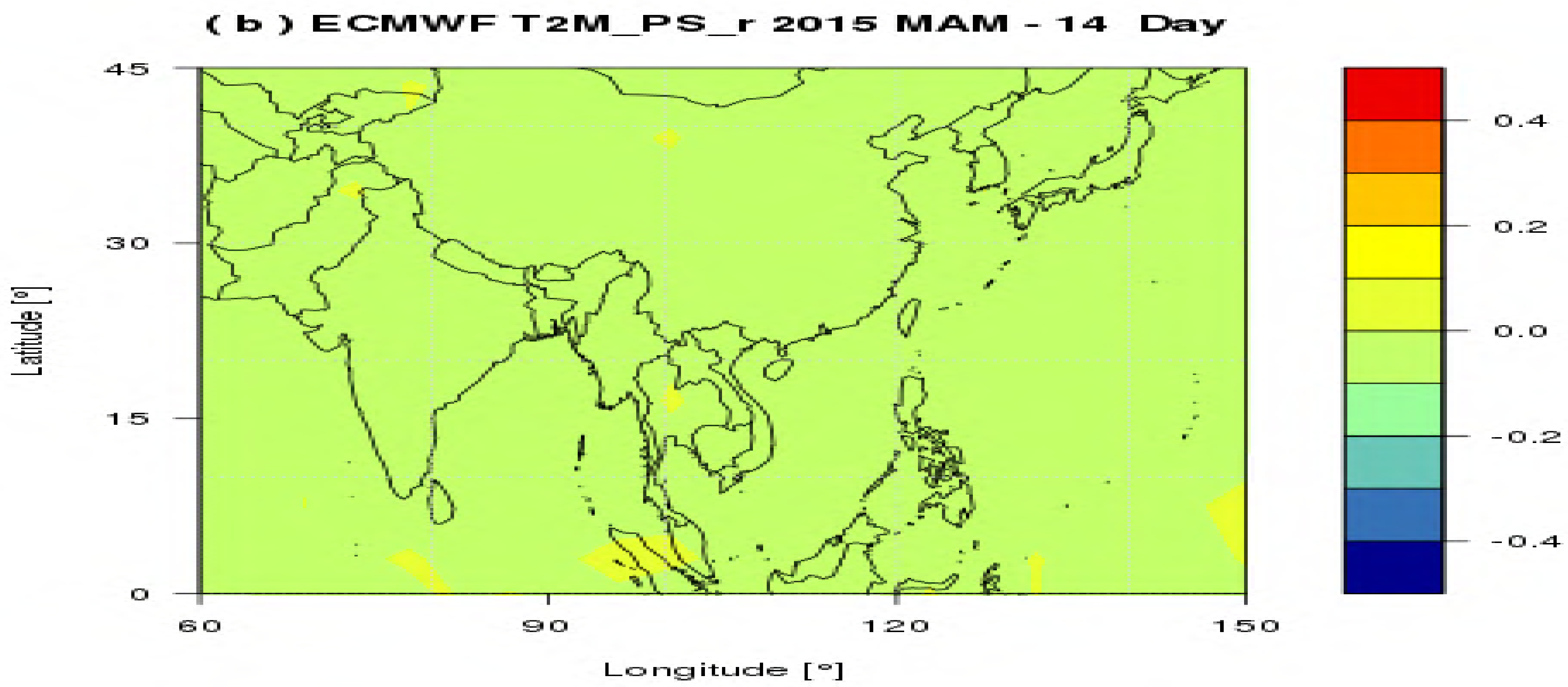
ECMWF S2S 2015 Spring T2M PS_ $\bar{p}_Q$(y)



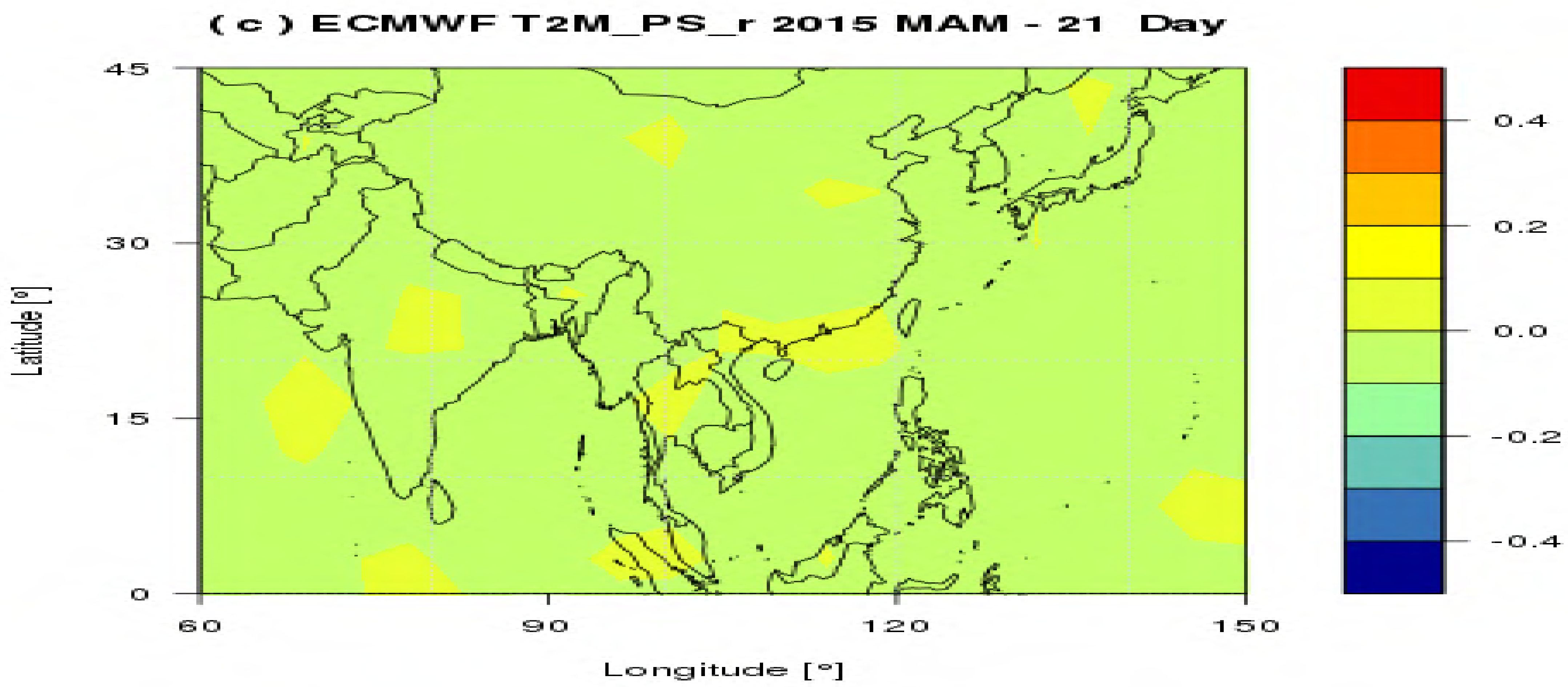
ECMWF S2S 2015 Spring T2M PS_ γ_{Q_p}



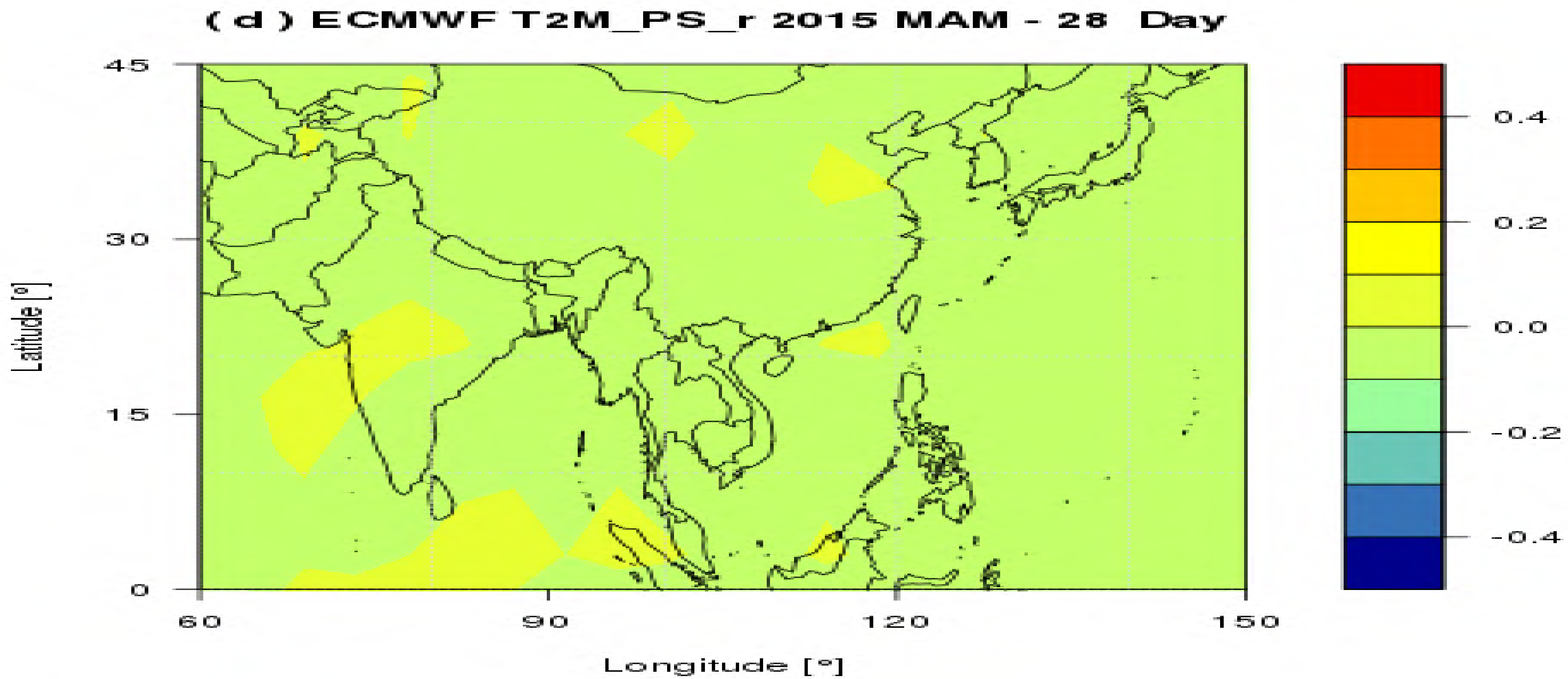
ECMWF S2S 2015 Spring T2M PS_ γ_{Q_p}



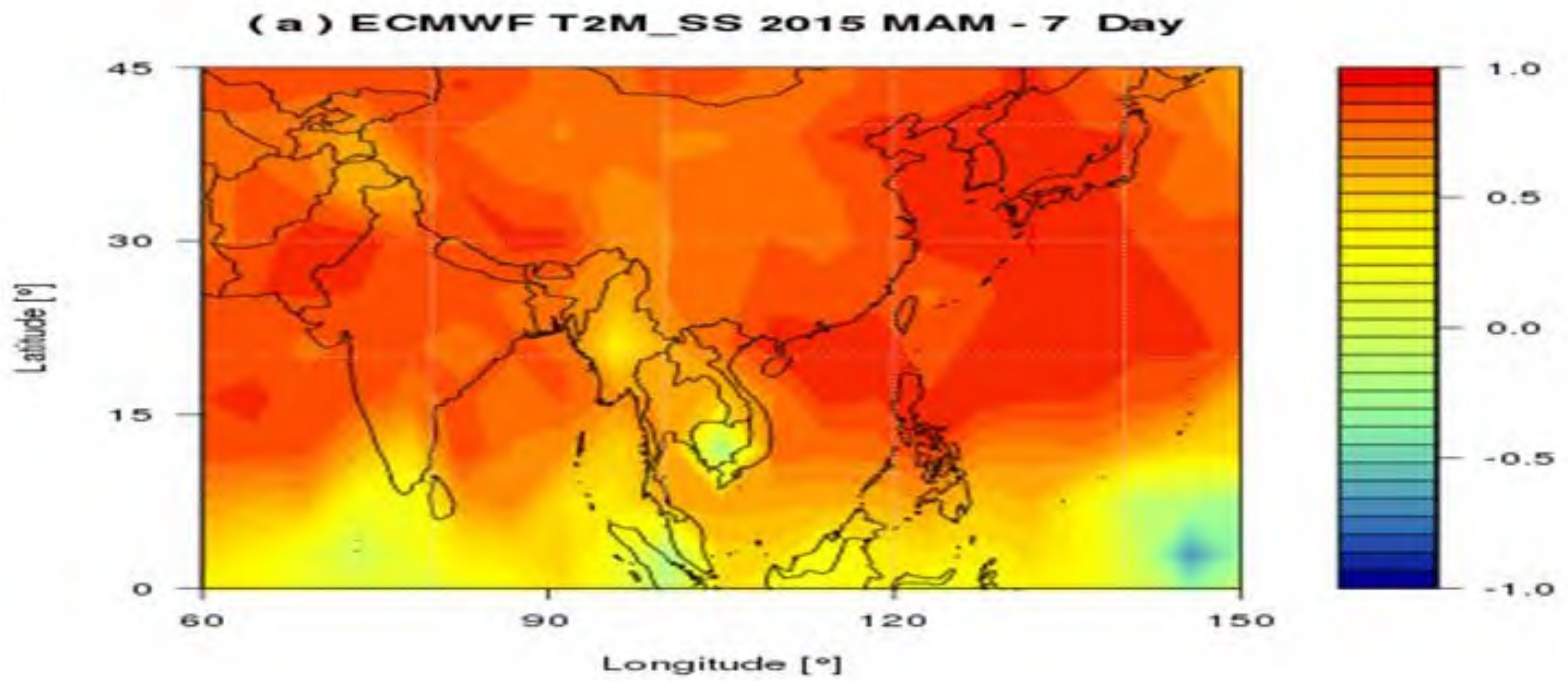
ECMWF S2S 2015 Spring T2M PS_ γ_{Q_p}



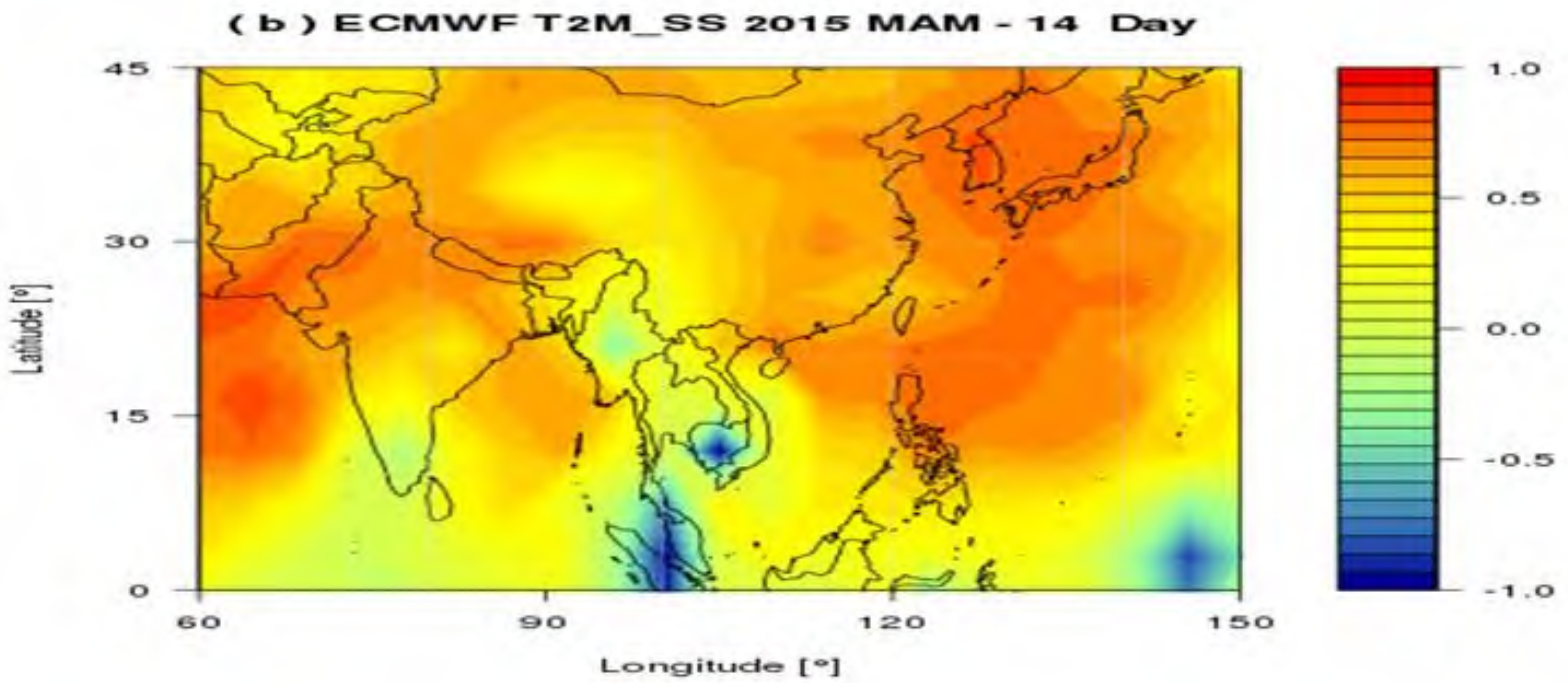
ECMWF S2S 2015 Spring T2M PS_ γ_{Q_p}



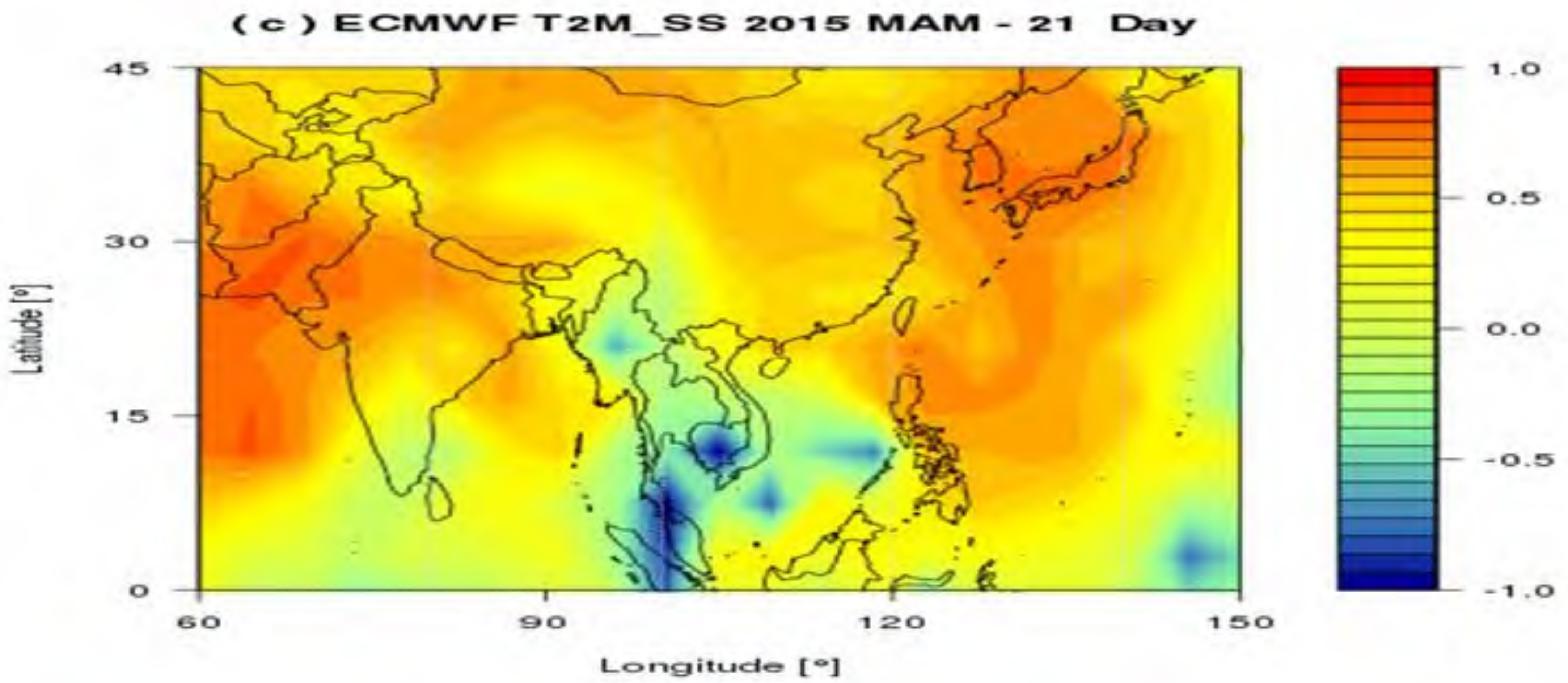
ECMWF S2S 2015 Spring T2M SS



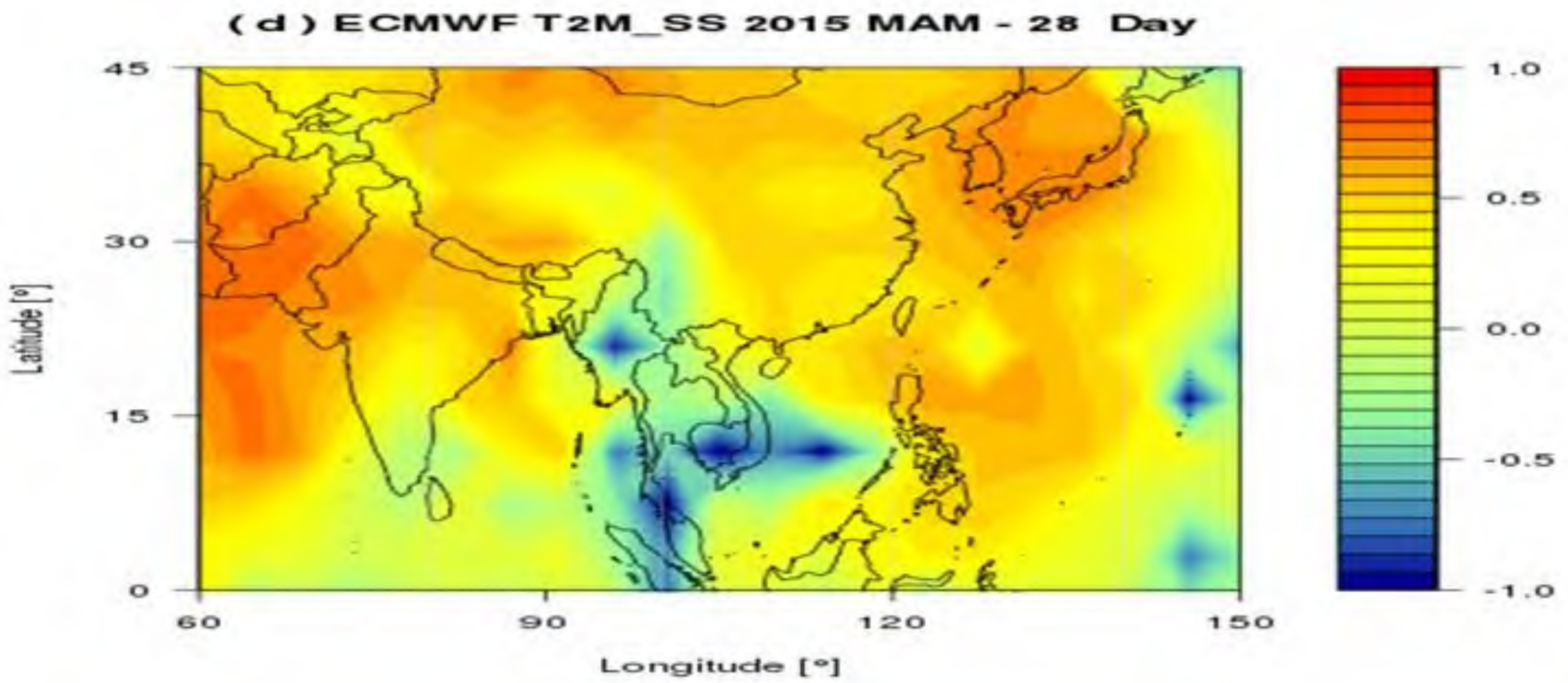
ECMWF S2S 2015 Spring T2M SS



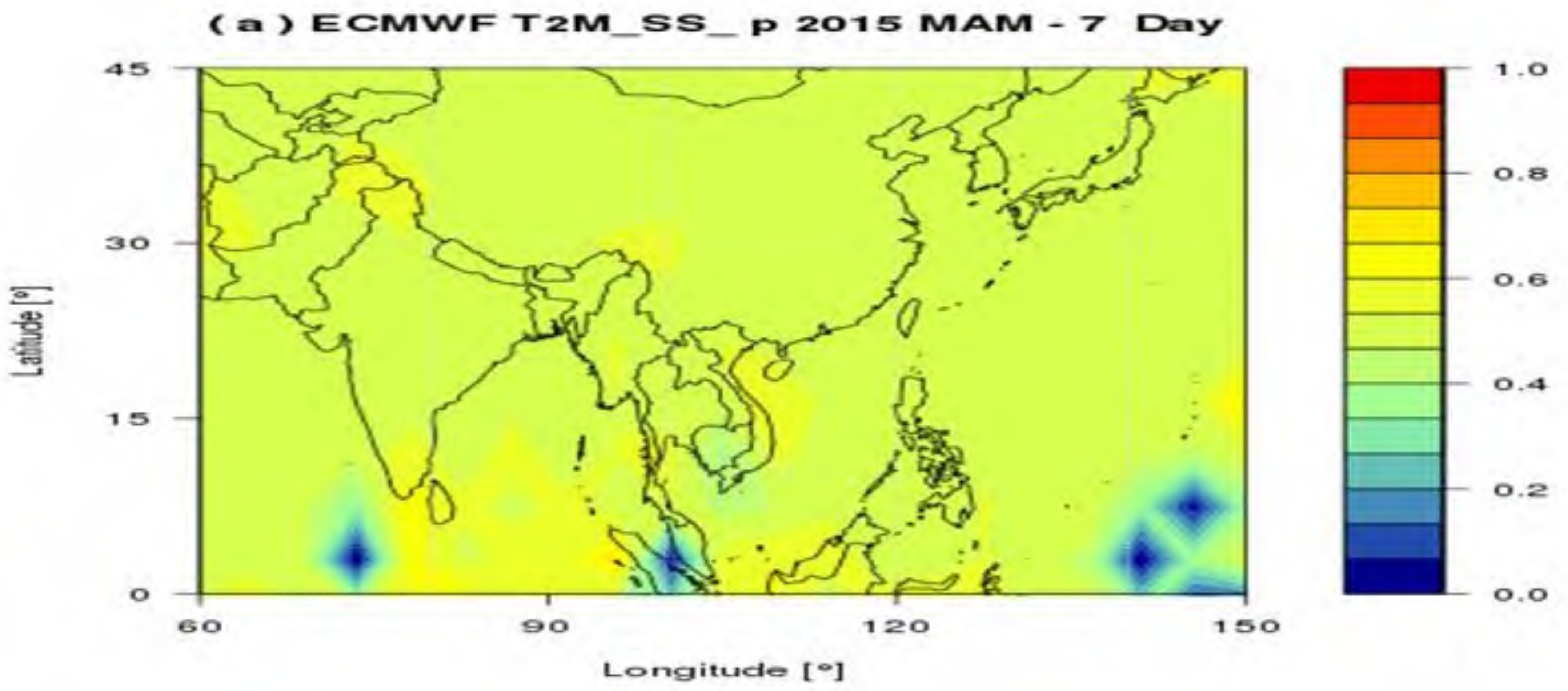
ECMWF S2S 2015 Spring T2M SS



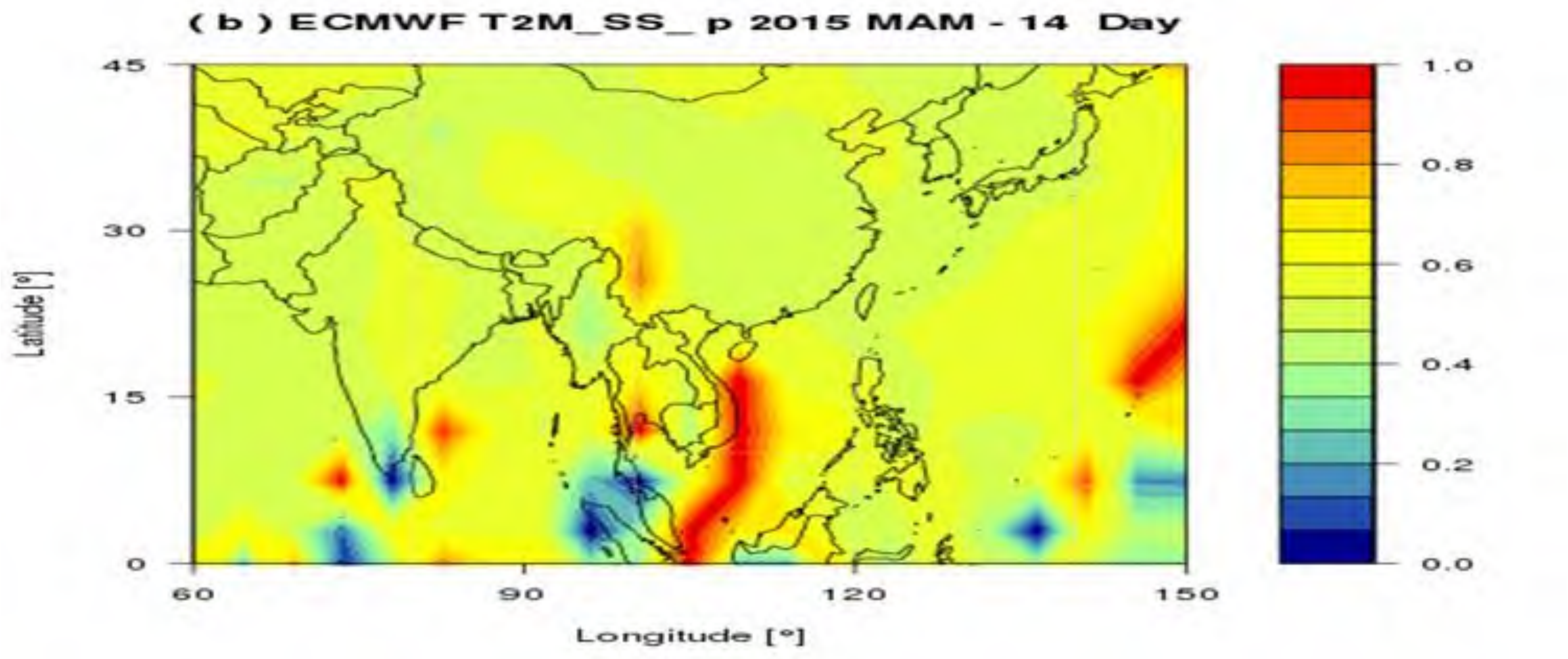
ECMWF S2S 2015 Spring T2M SS



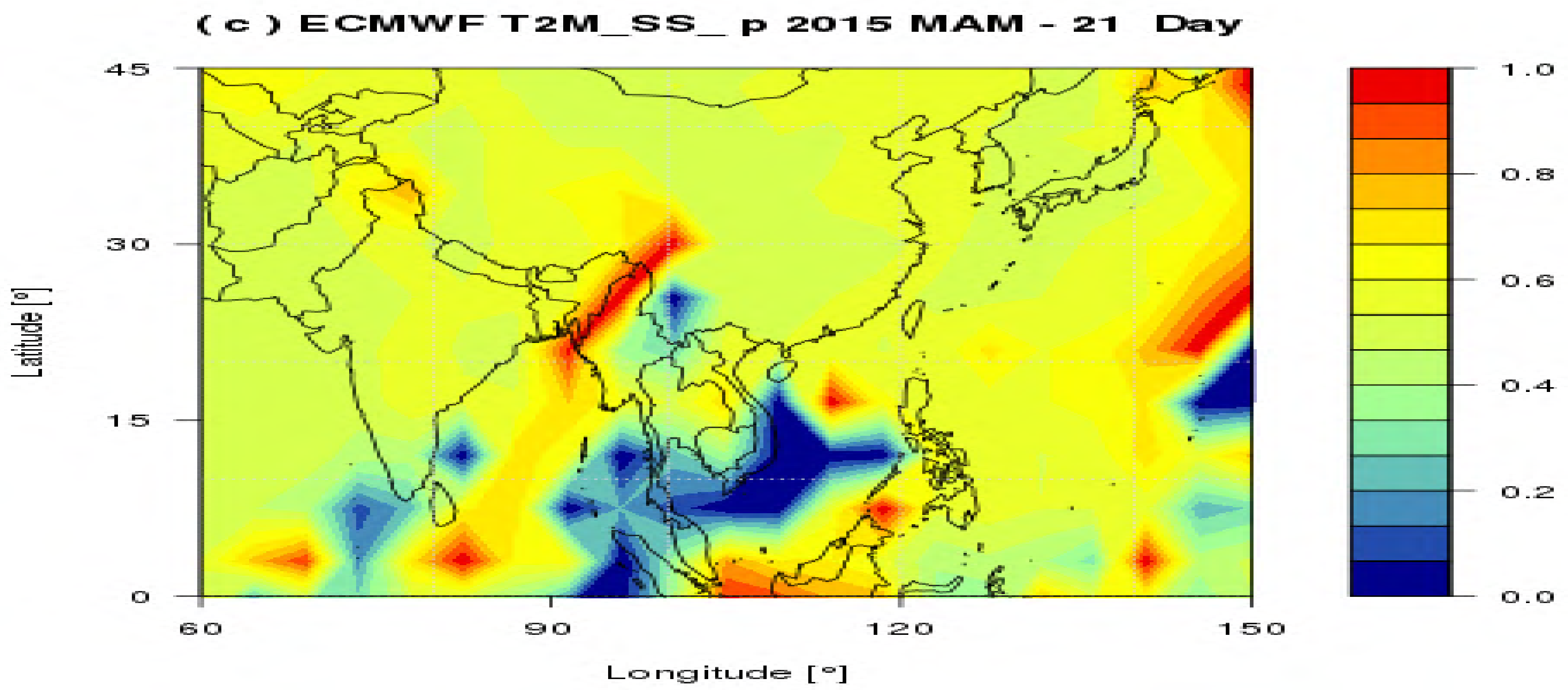
ECMWF S2S 2015 Spring T2M SS_ $\bar{p}_Q$



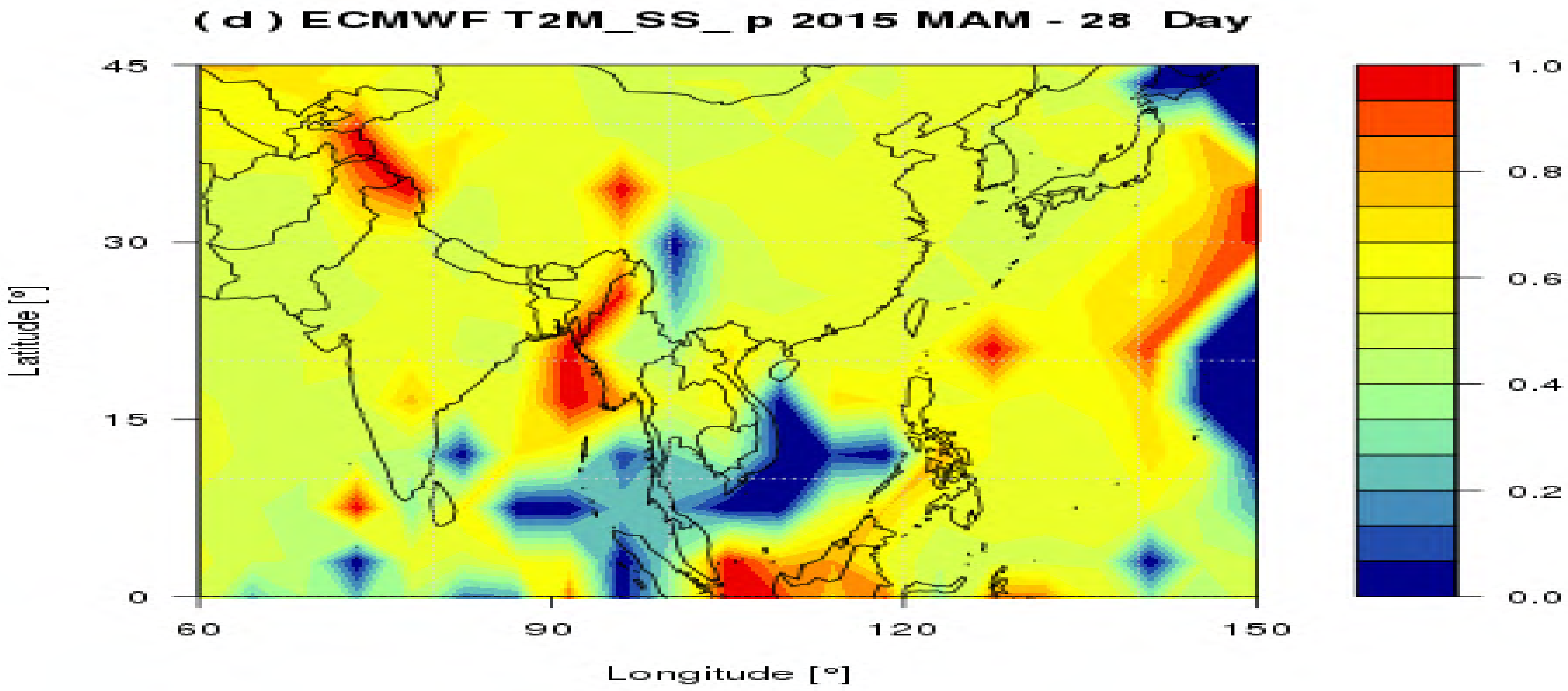
ECMWF S2S 2015 Spring T2M SS_ $\bar{p}_Q$



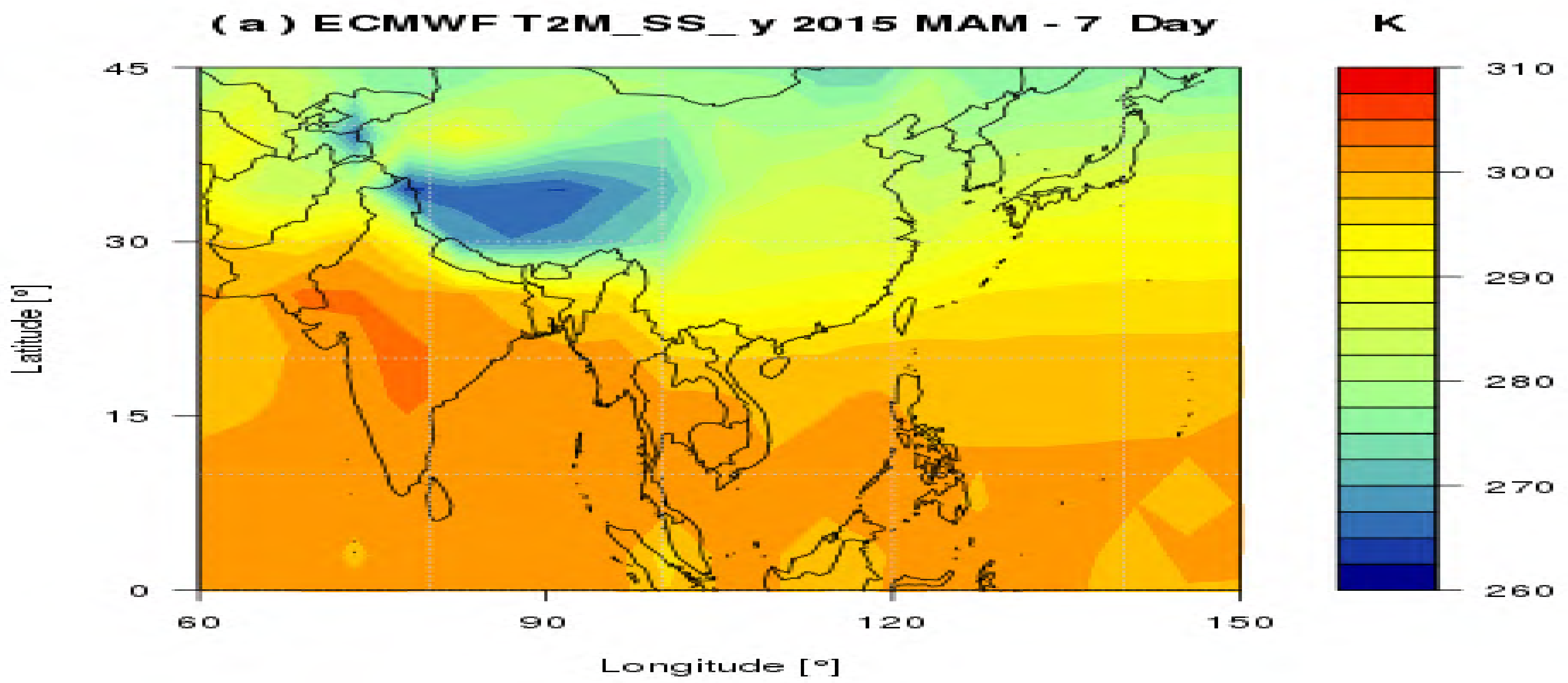
ECMWF S2S 2015 Spring T2M SS_ $\bar{p}_Q$



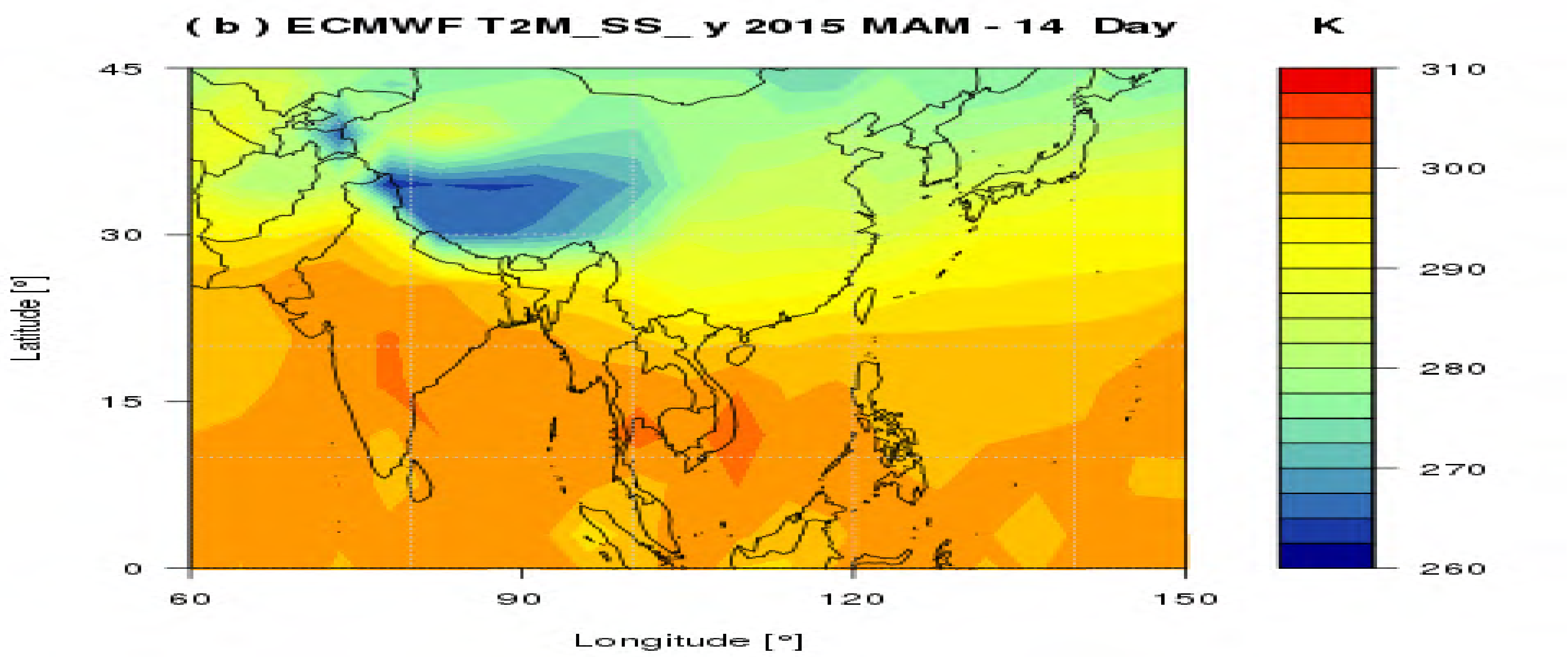
ECMWF S2S 2015 Spring T2M SS_ $\bar{p}_Q$



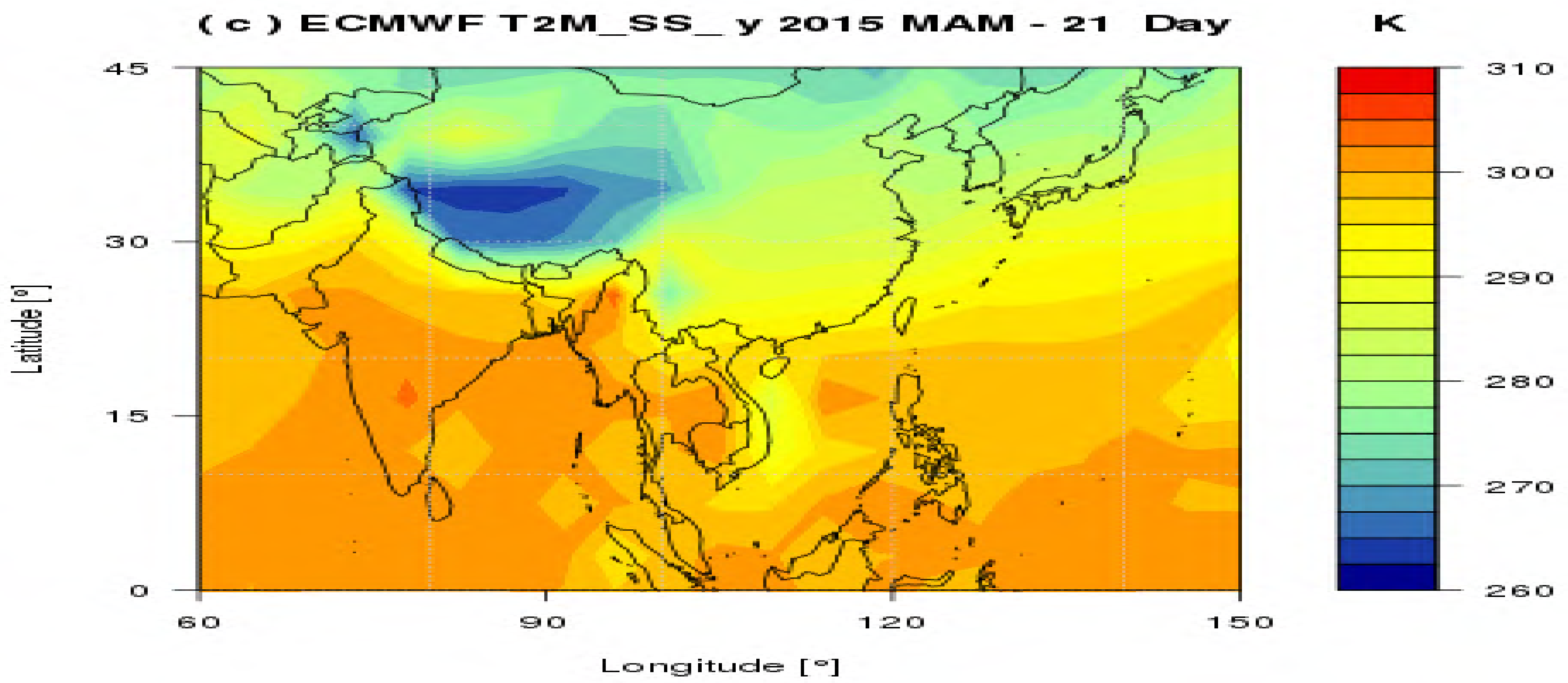
ECMWF S2S 2015 Spring T2M SS_ $\bar{p}_Q$(y)



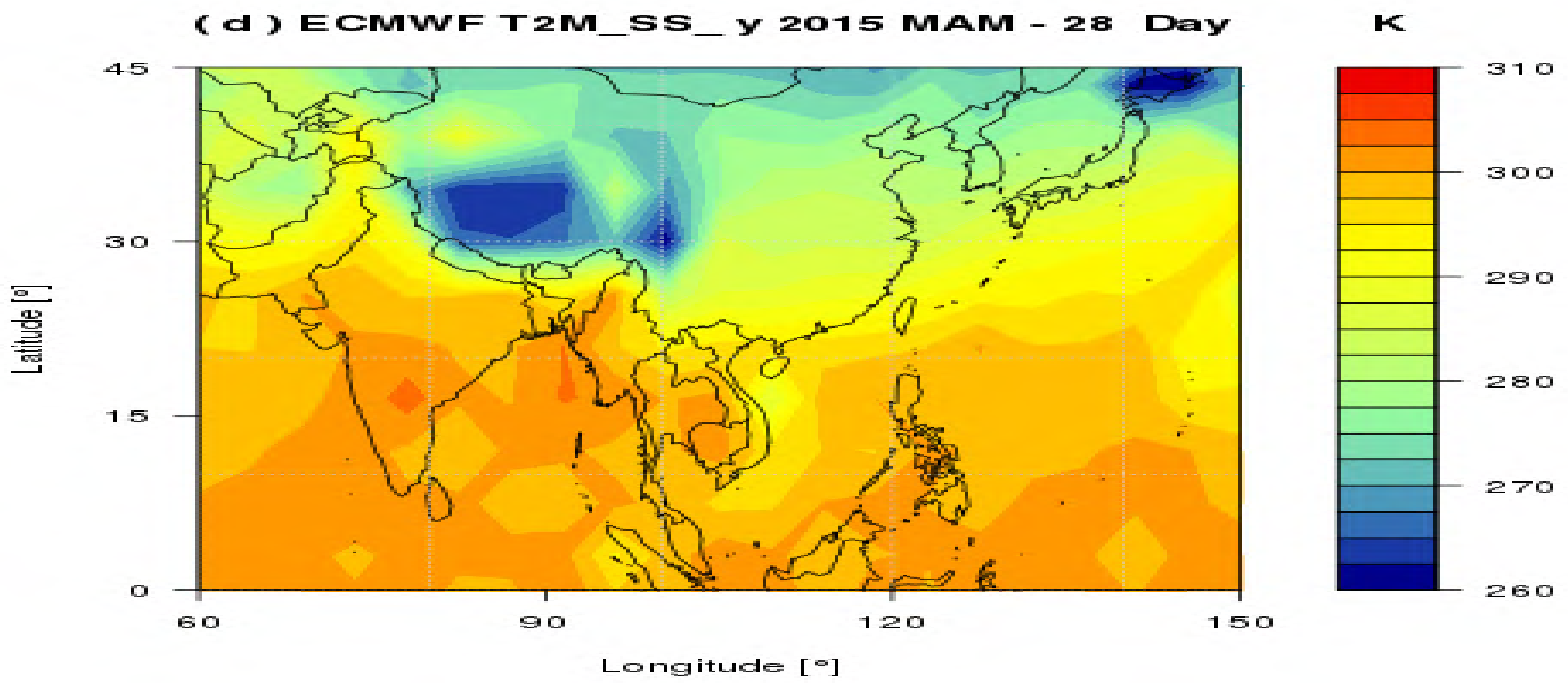
ECMWF S2S 2015 Spring T2M SS_ $\bar{p}_Q$(y)



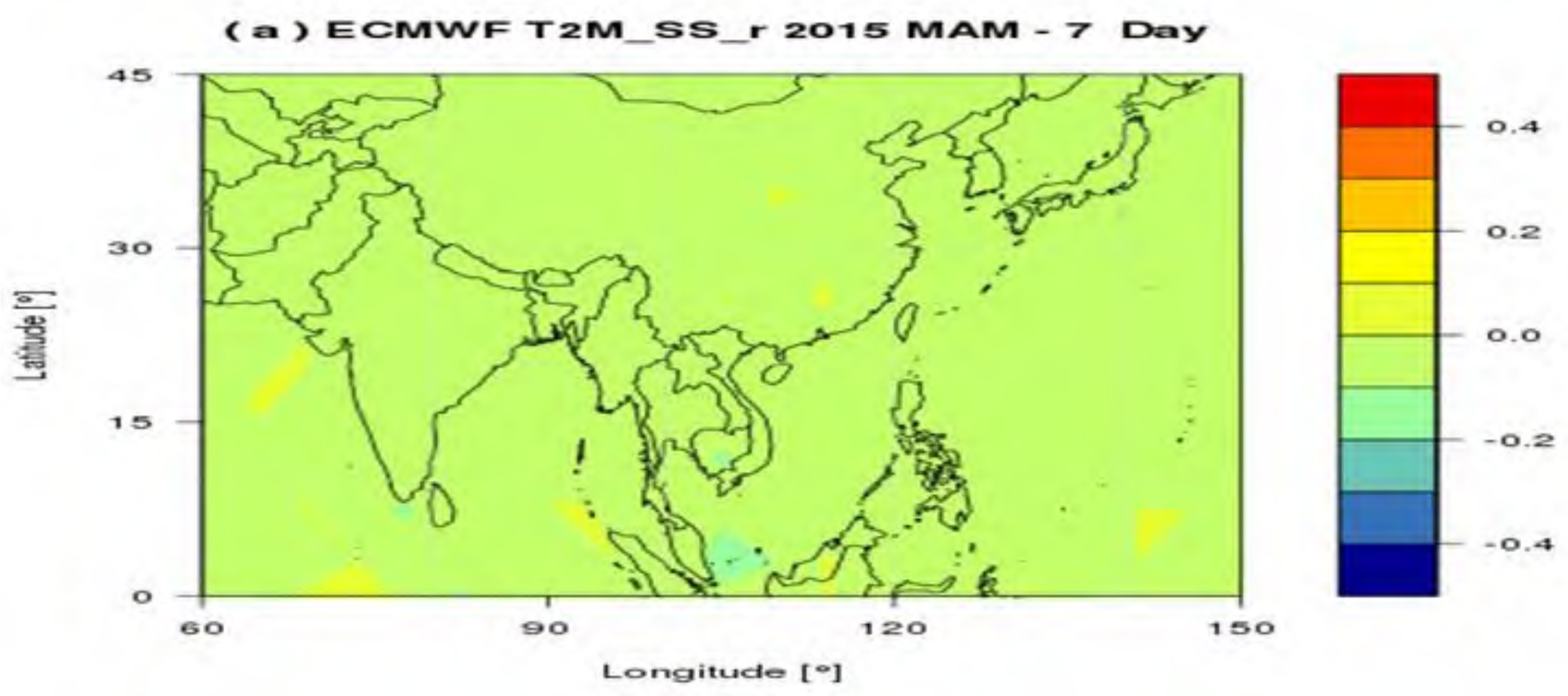
ECMWF S2S 2015 Spring T2M SS_ $\bar{p}_Q$(y)



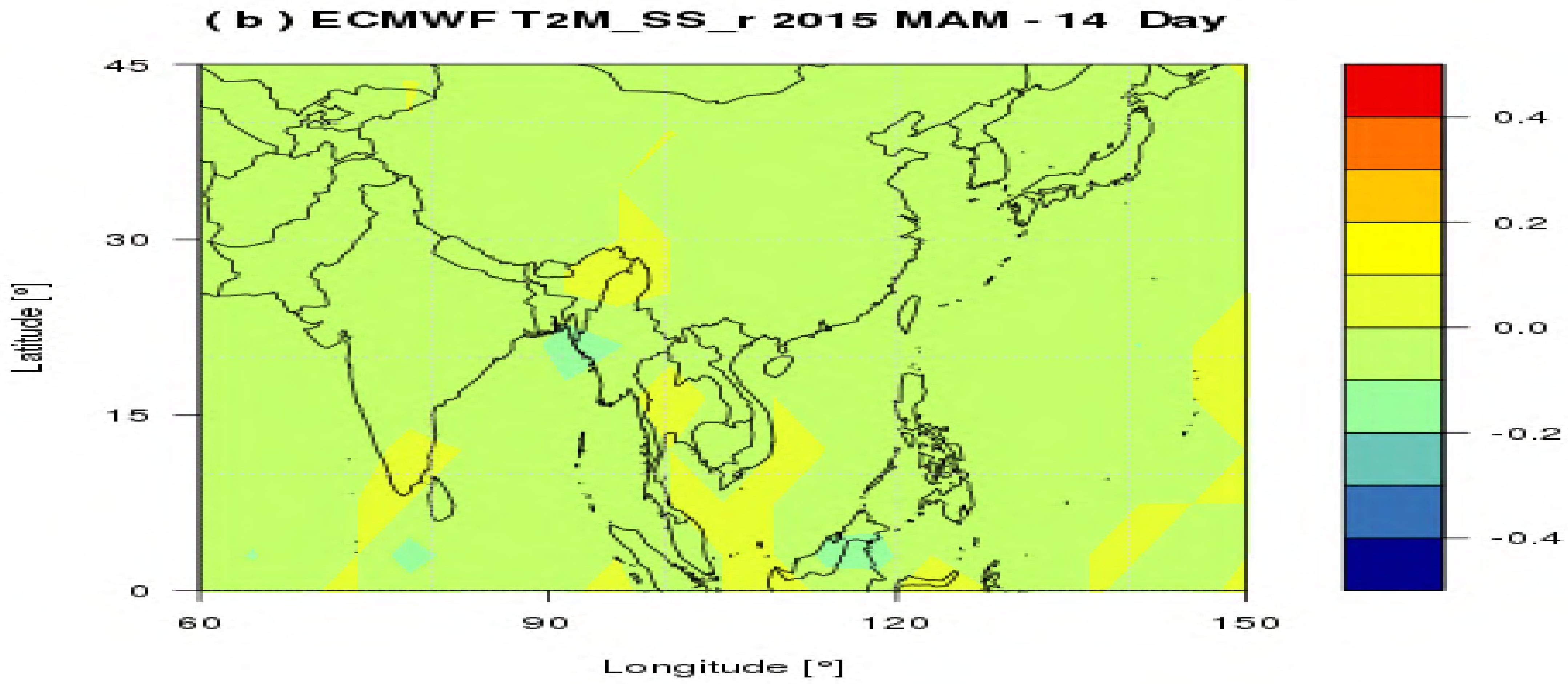
ECMWF S2S 2015 Spring T2M SS_ $\bar{p}_Q$(y)



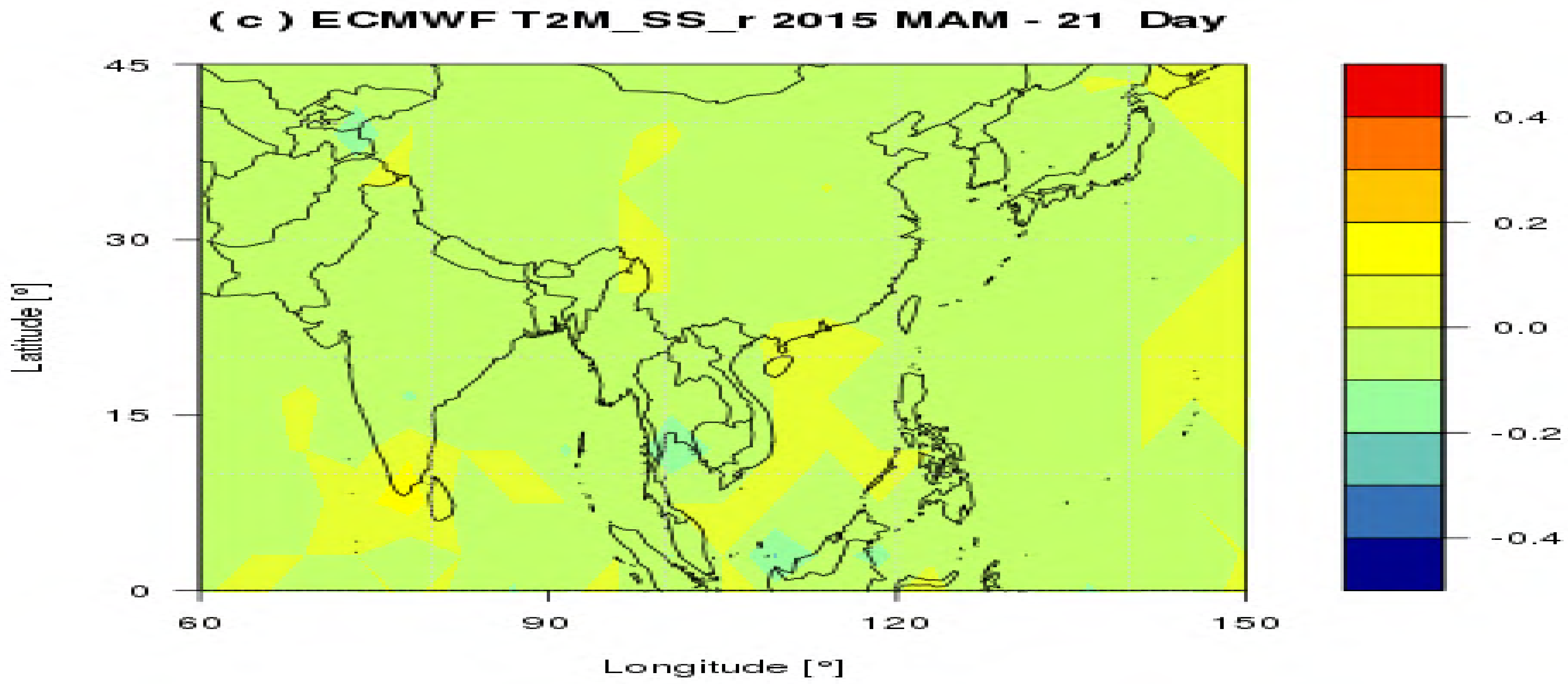
ECMWF S2S 2015 Spring T2M SS_ γ_{Q_p}



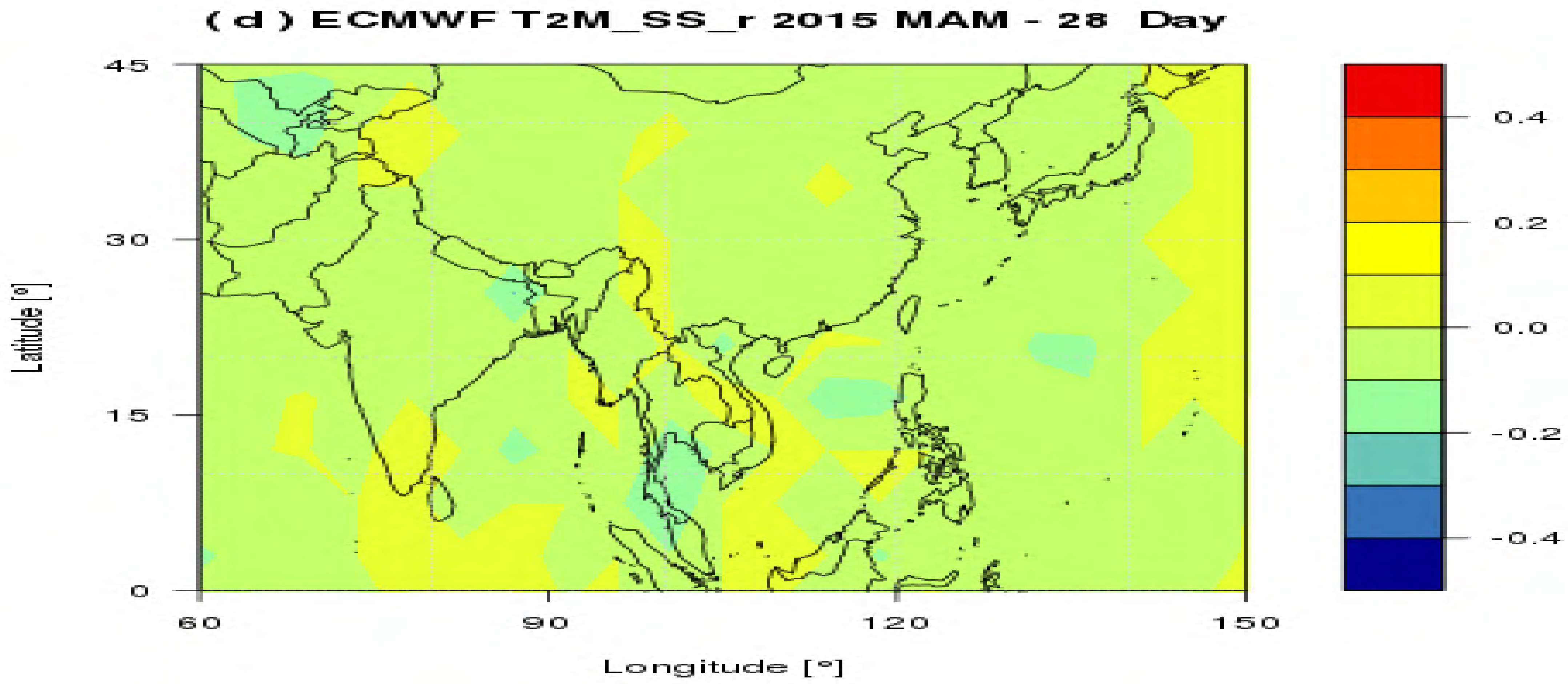
ECMWF S2S 2015 Spring T2M SS_ γ_{Q_p}



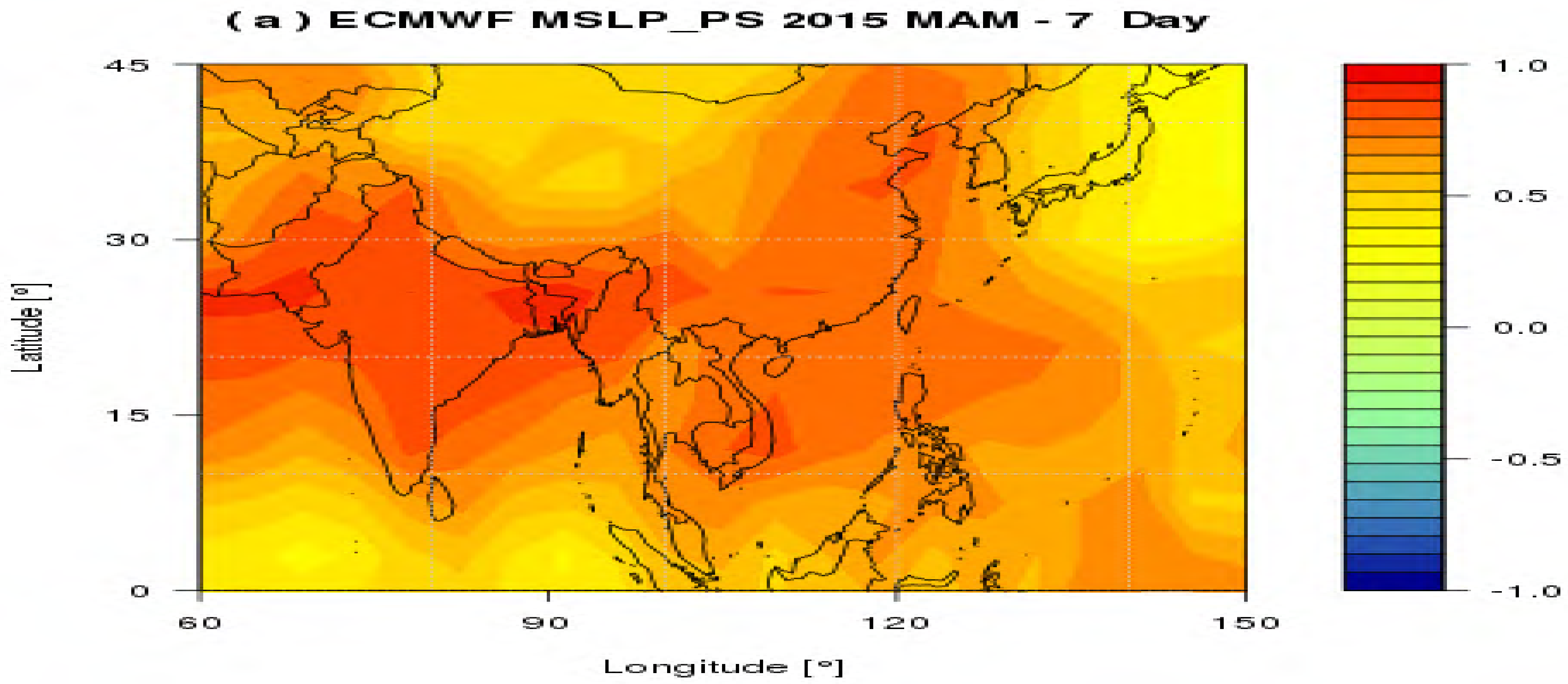
ECMWF S2S 2015 Spring T2M SS_ γ_{Q_p}



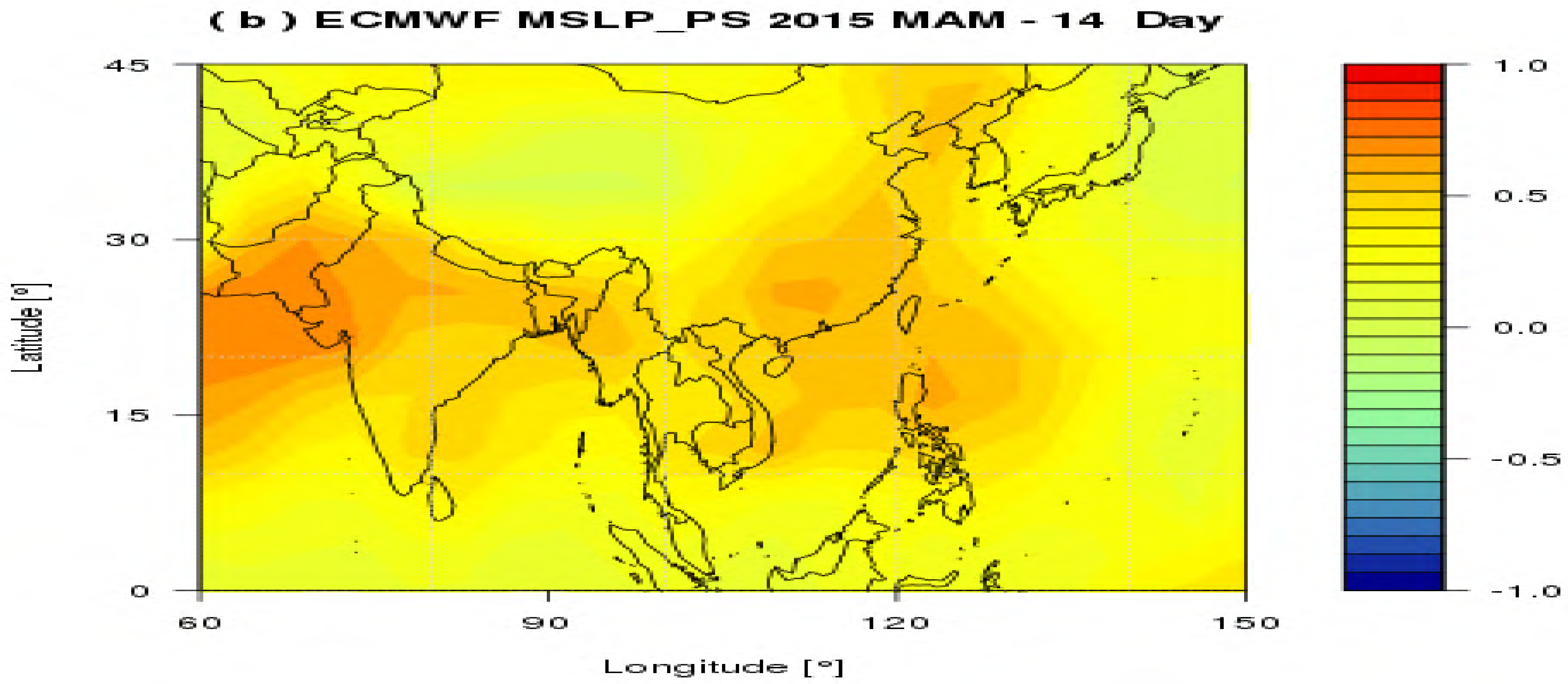
ECMWF S2S 2015 Spring T2M SS_ γ_{Q_p}



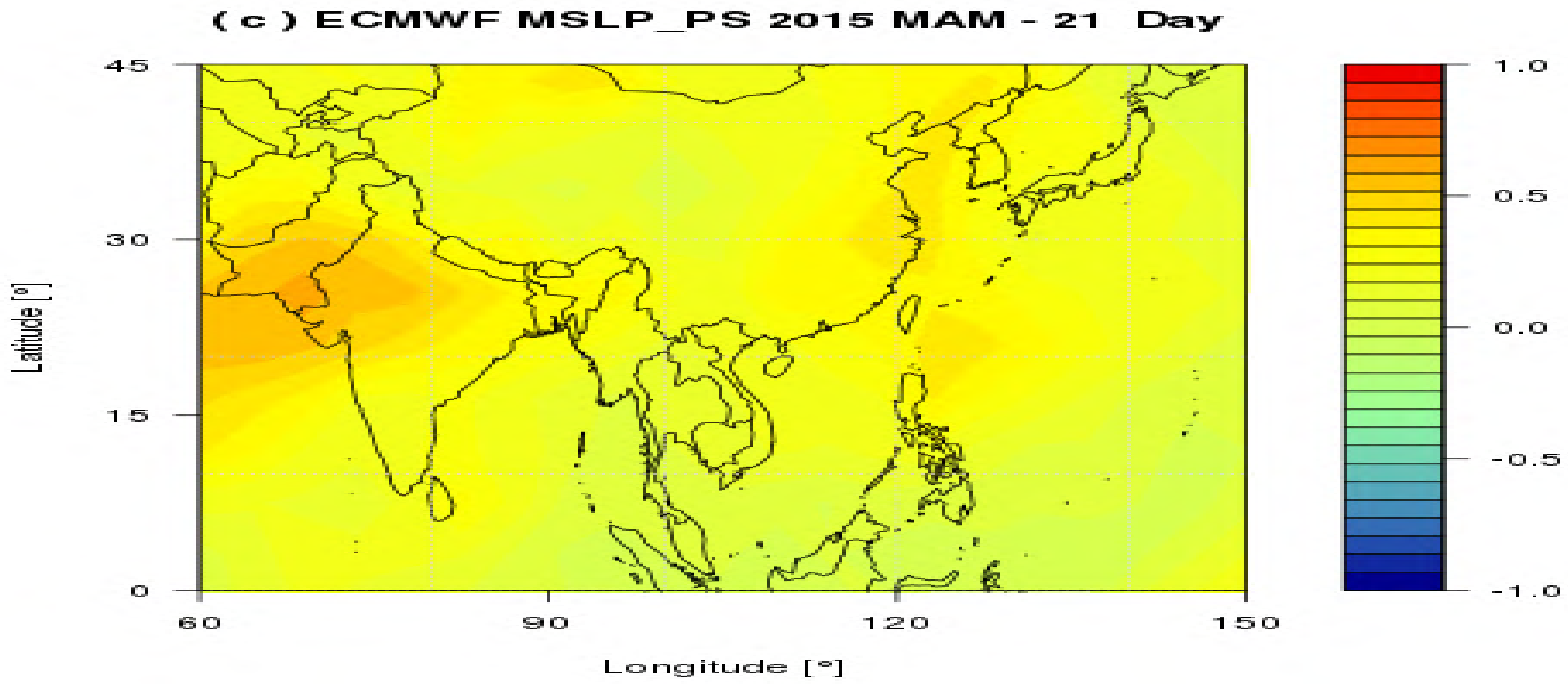
ECMWF S2S 2015 Spring MSLP PS



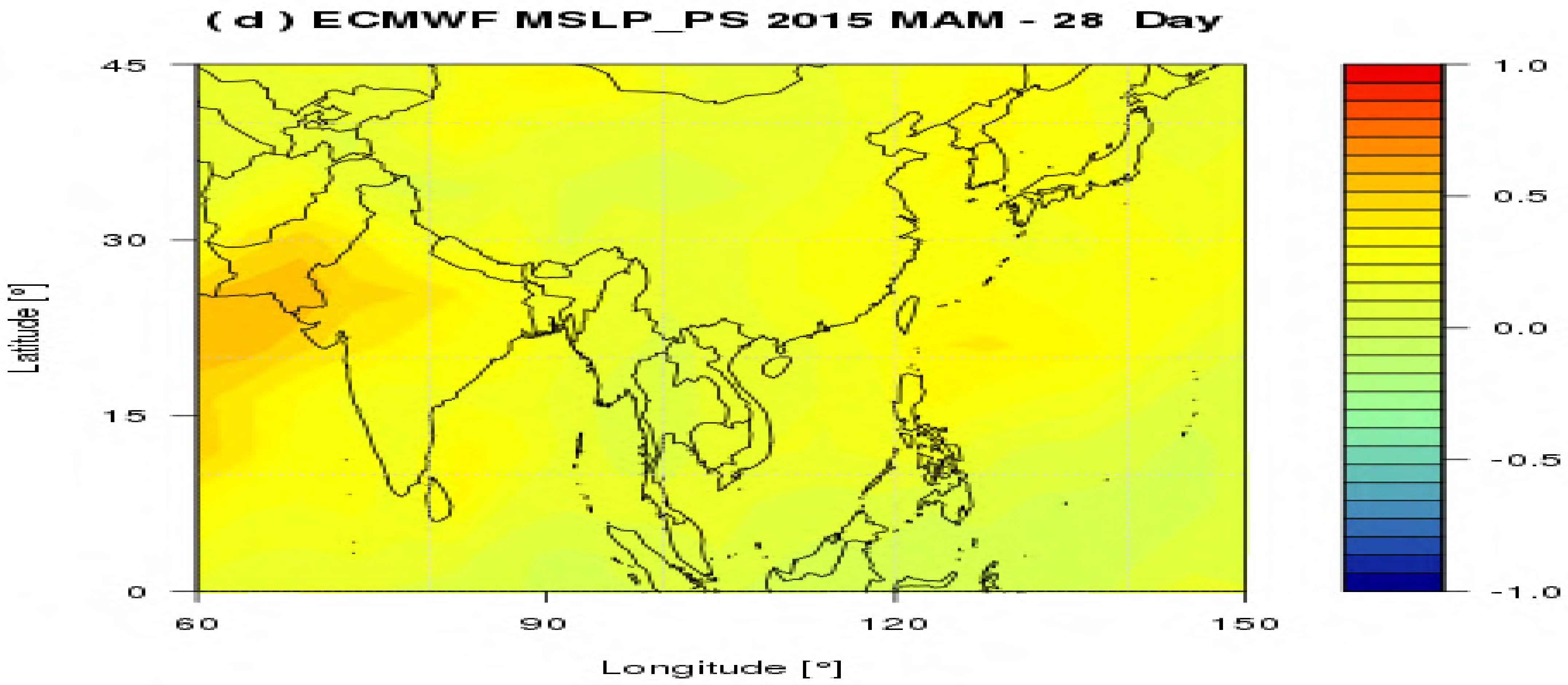
ECMWF S2S 2015 Spring MSLP PS



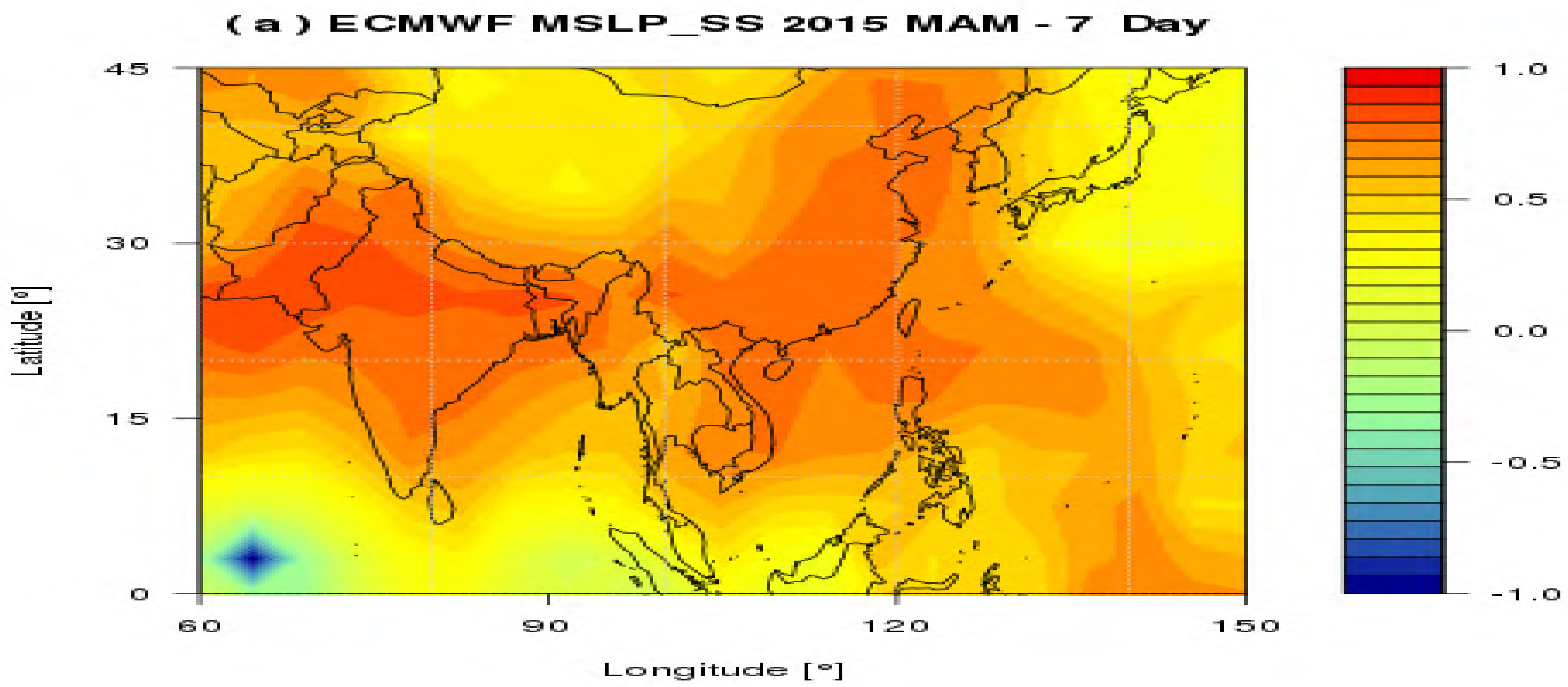
ECMWF S2S 2015 Spring MSLP PS



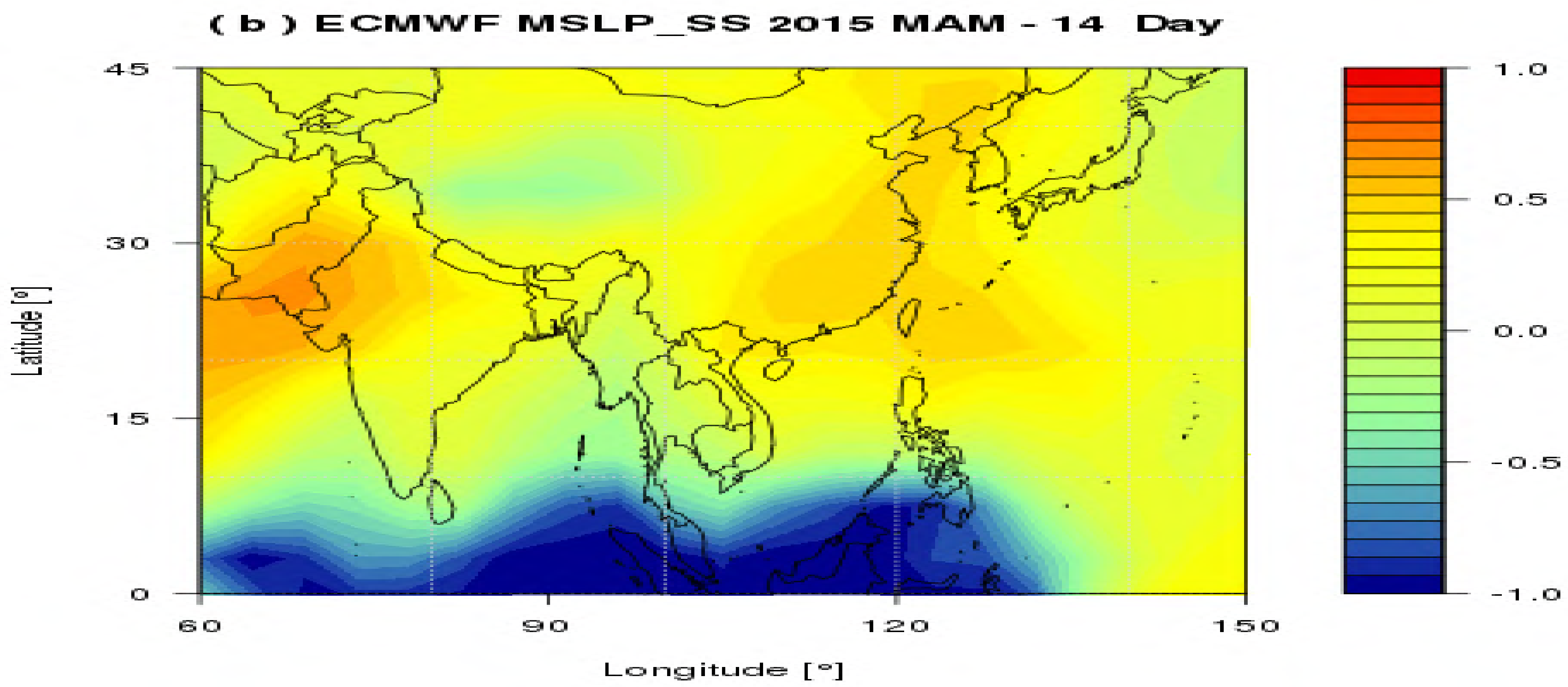
ECMWF S2S 2015 Spring MSLP PS



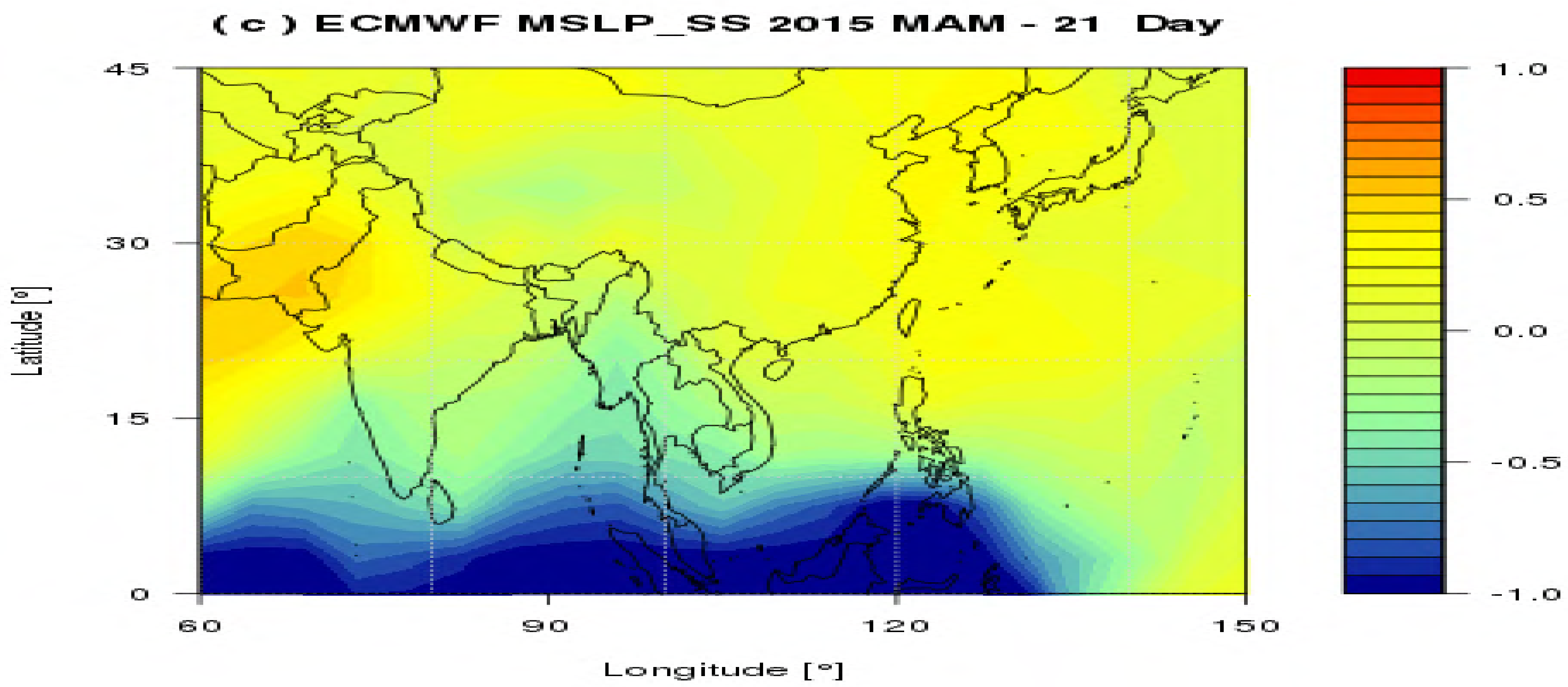
ECMWF S2S 2015 Spring MSLP SS



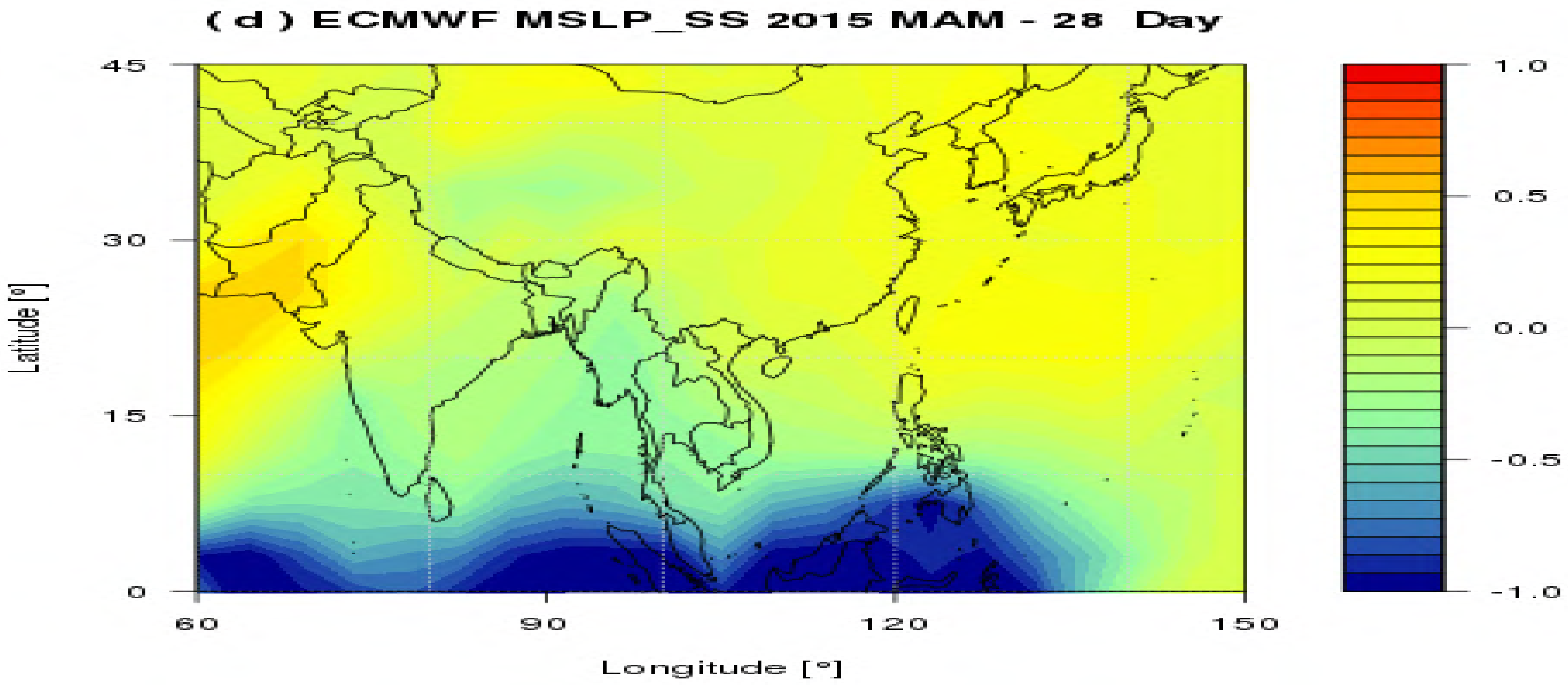
ECMWF S2S 2015 Spring MSLP SS



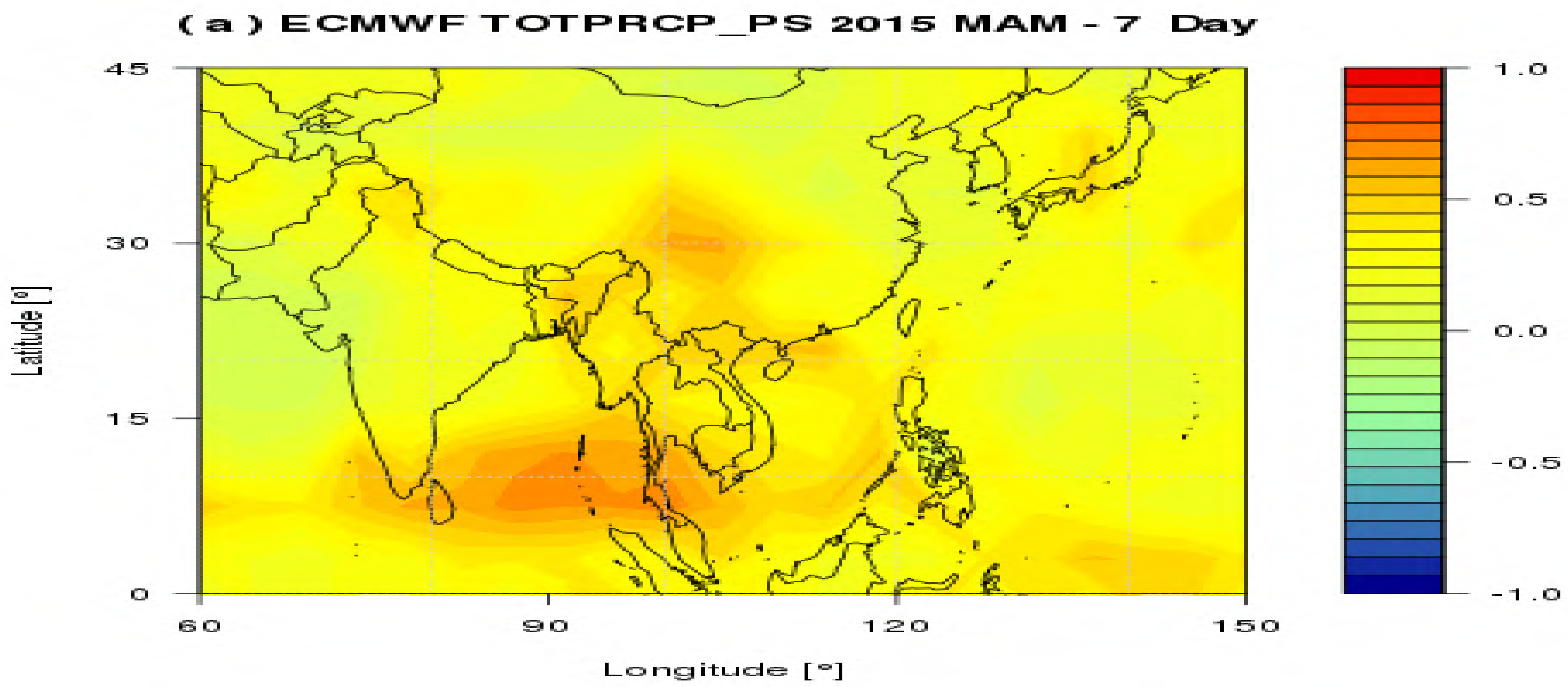
ECMWF S2S 2015 Spring MSLP SS



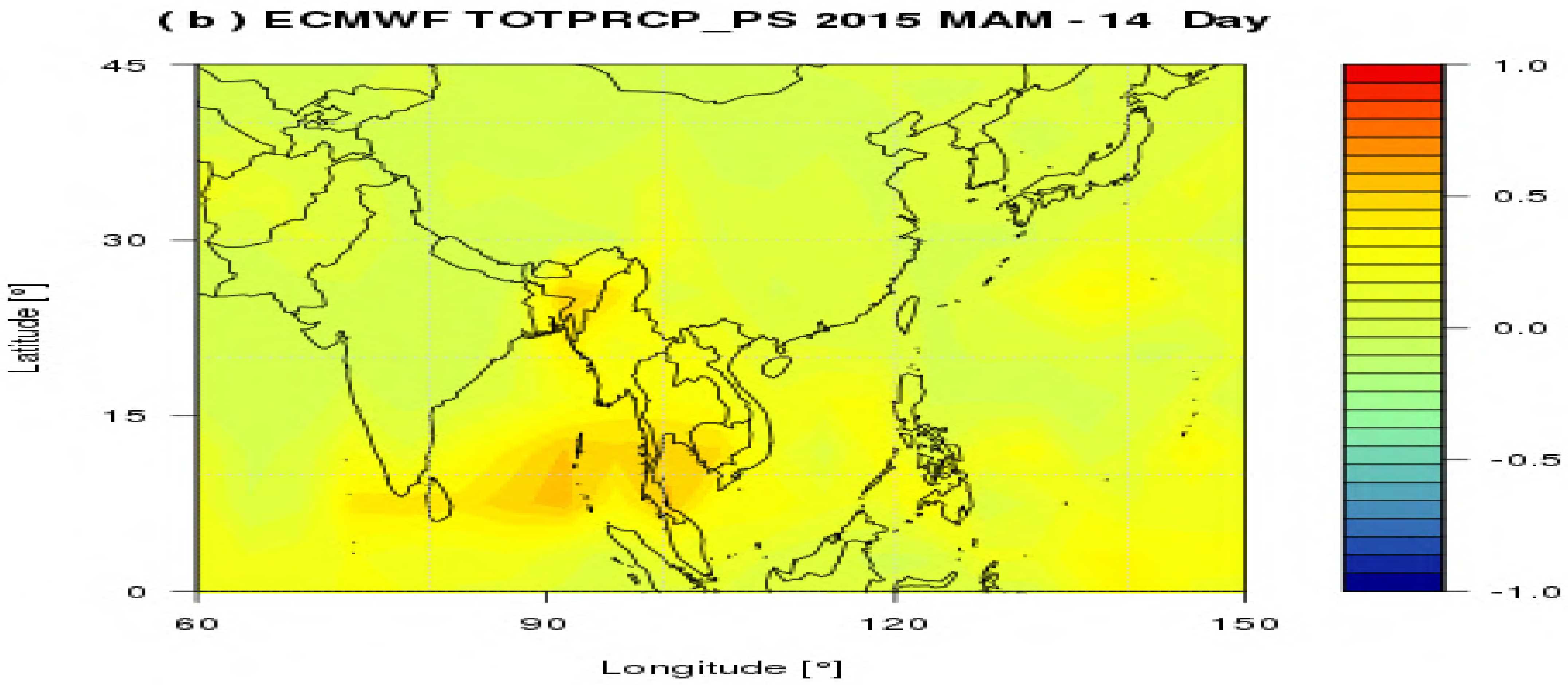
ECMWF S2S 2015 Spring MSLP SS



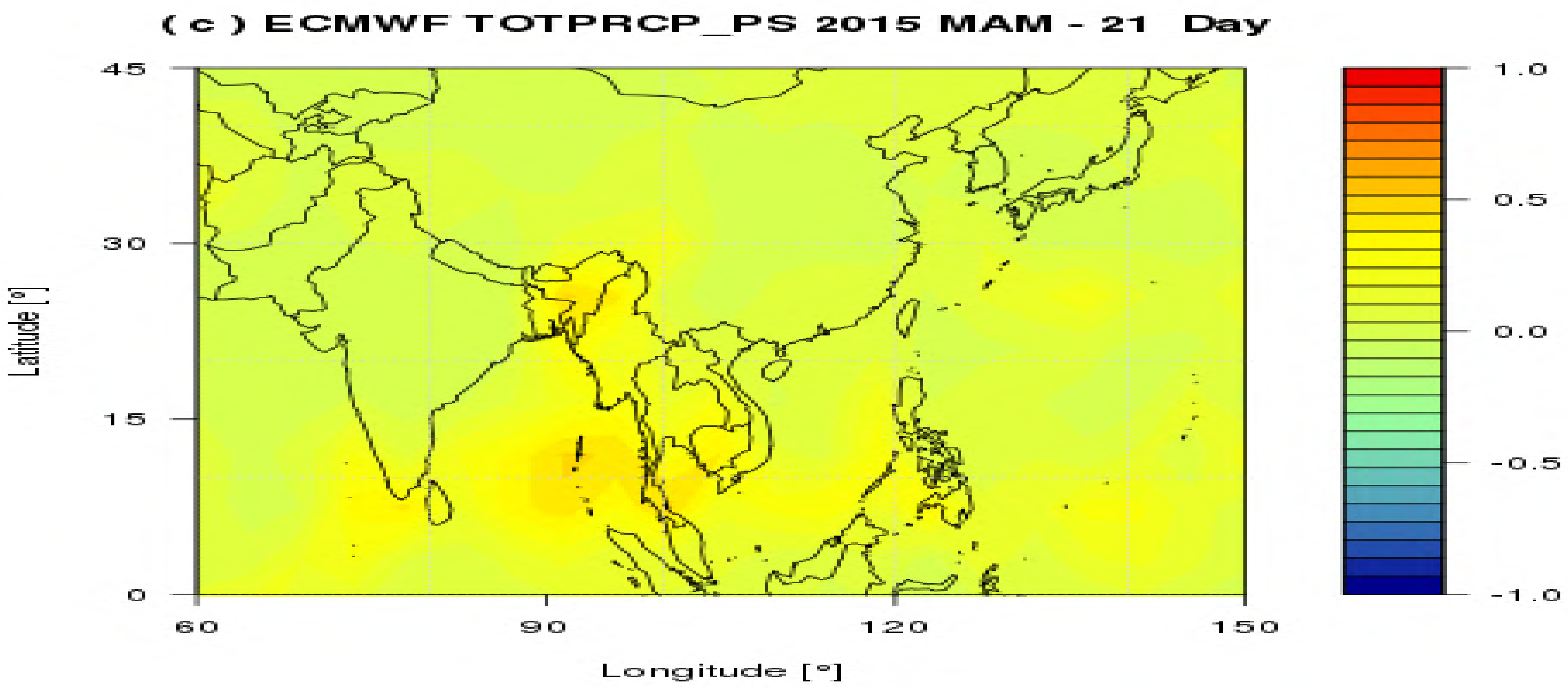
ECMWF S2S 2015 Spring TOTPRCP PS



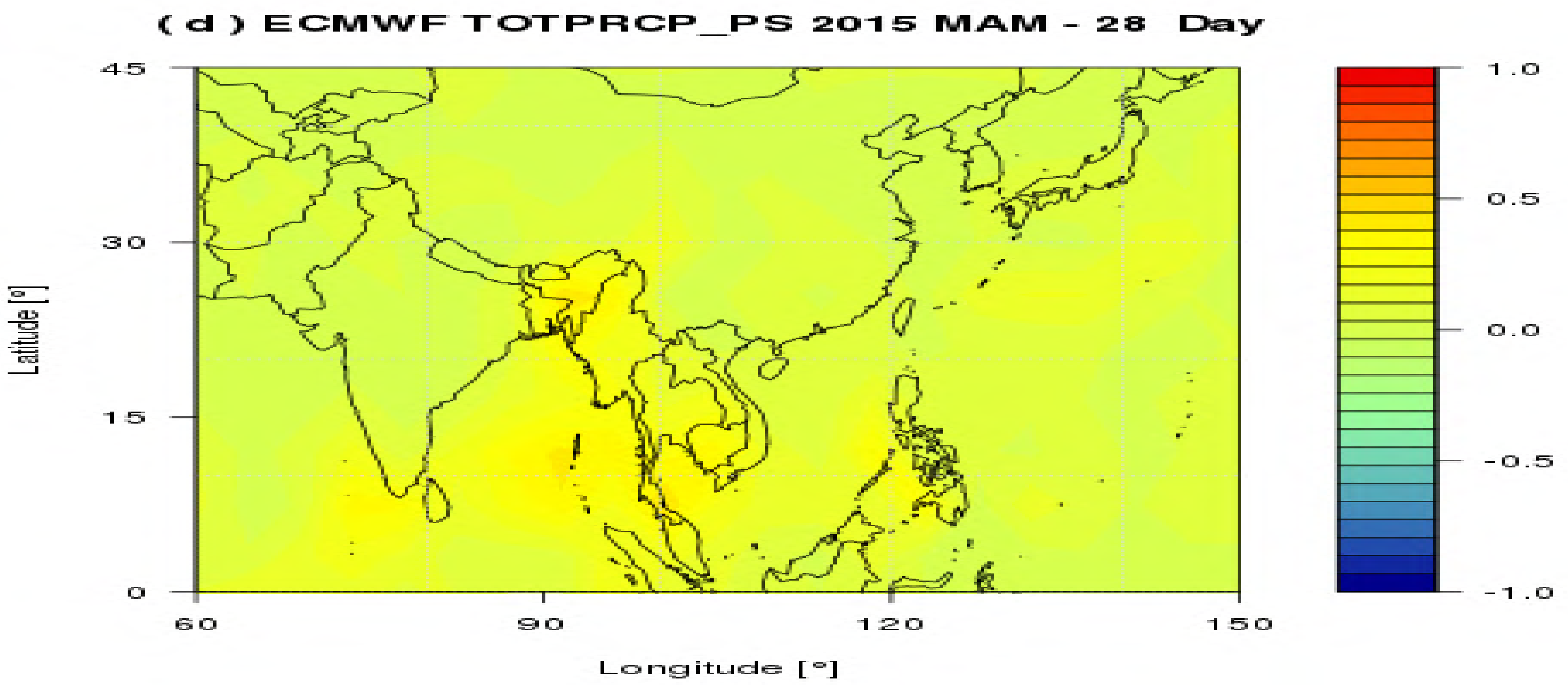
ECMWF S2S 2015 Spring TOTPRCP PS



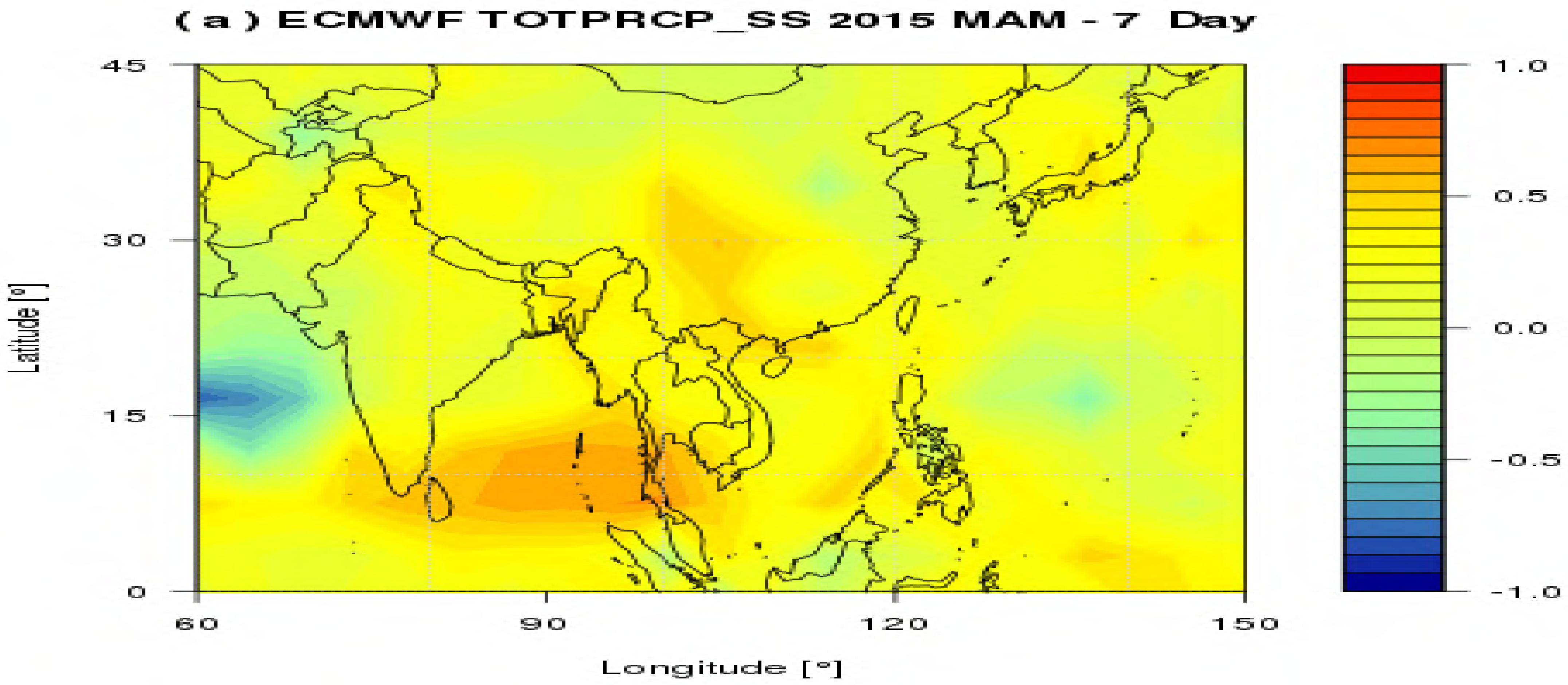
ECMWF S2S 2015 Spring TOTPRCP PS



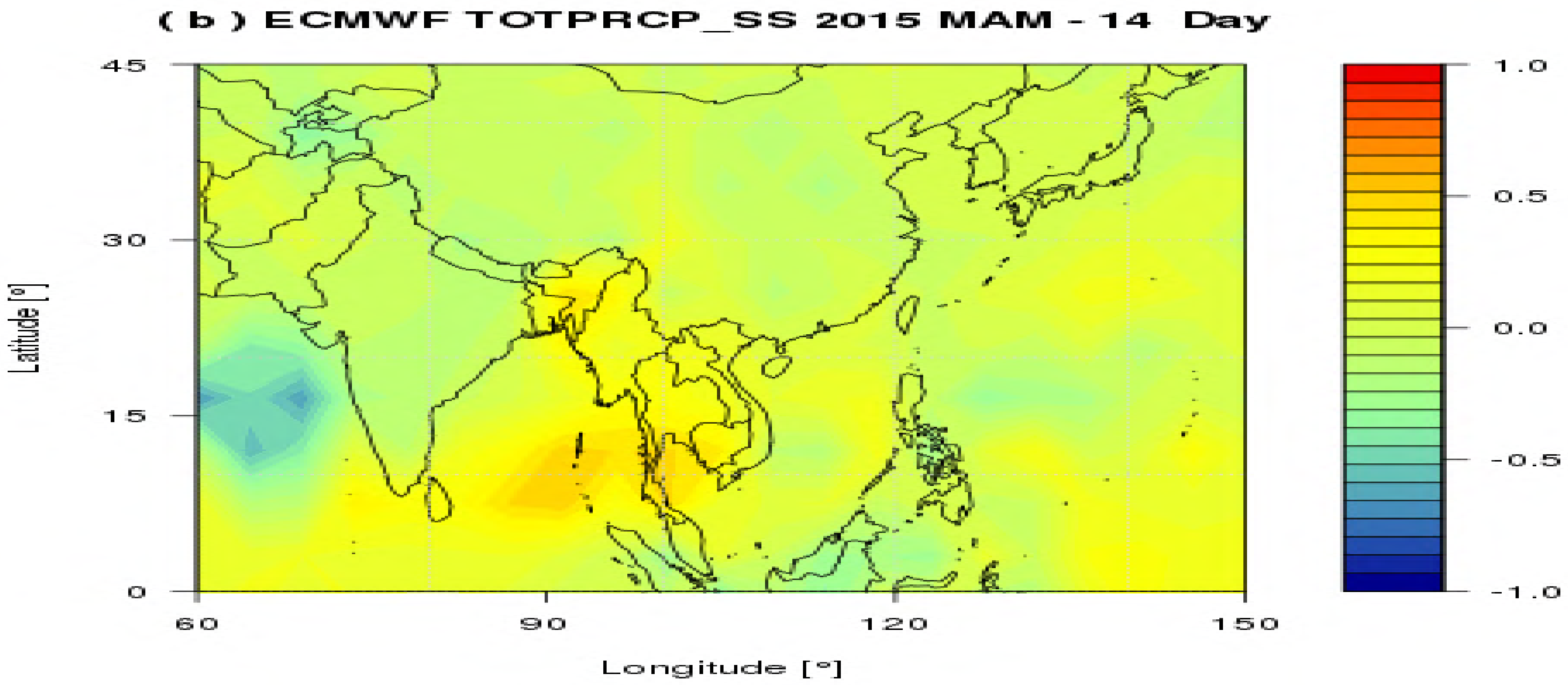
ECMWF S2S 2015 Spring TOTPRCP PS



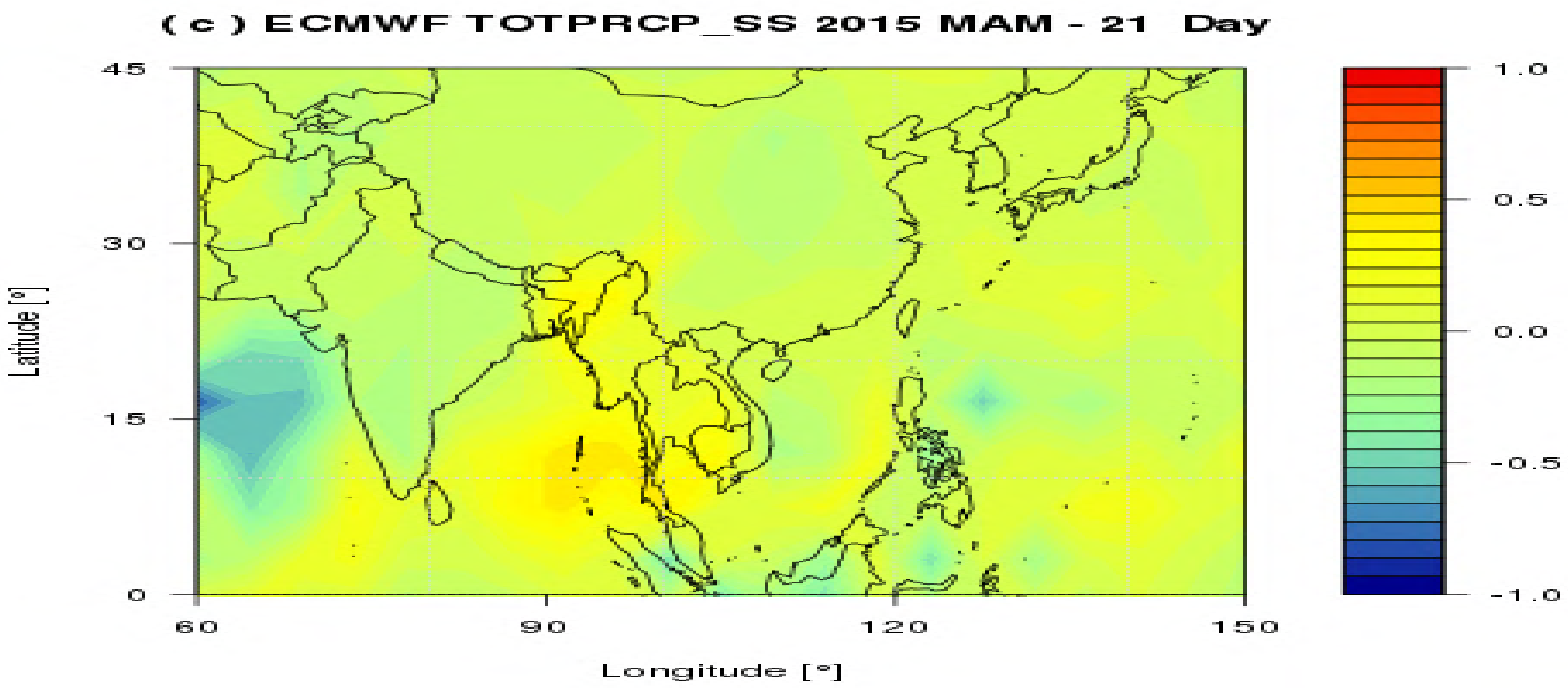
ECMWF S2S 2015 Spring TOTPRCP SS



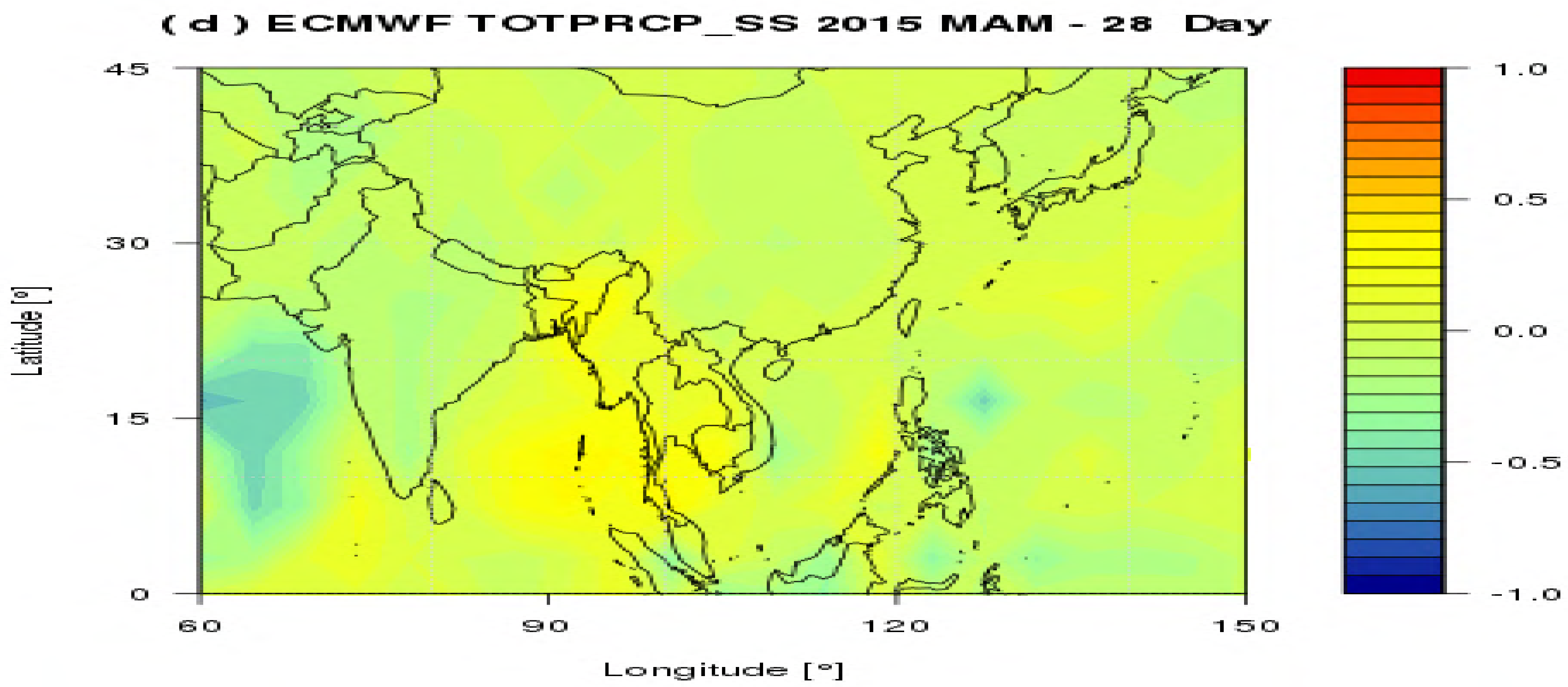
ECMWF S2S 2015 Spring TOTPRCP SS



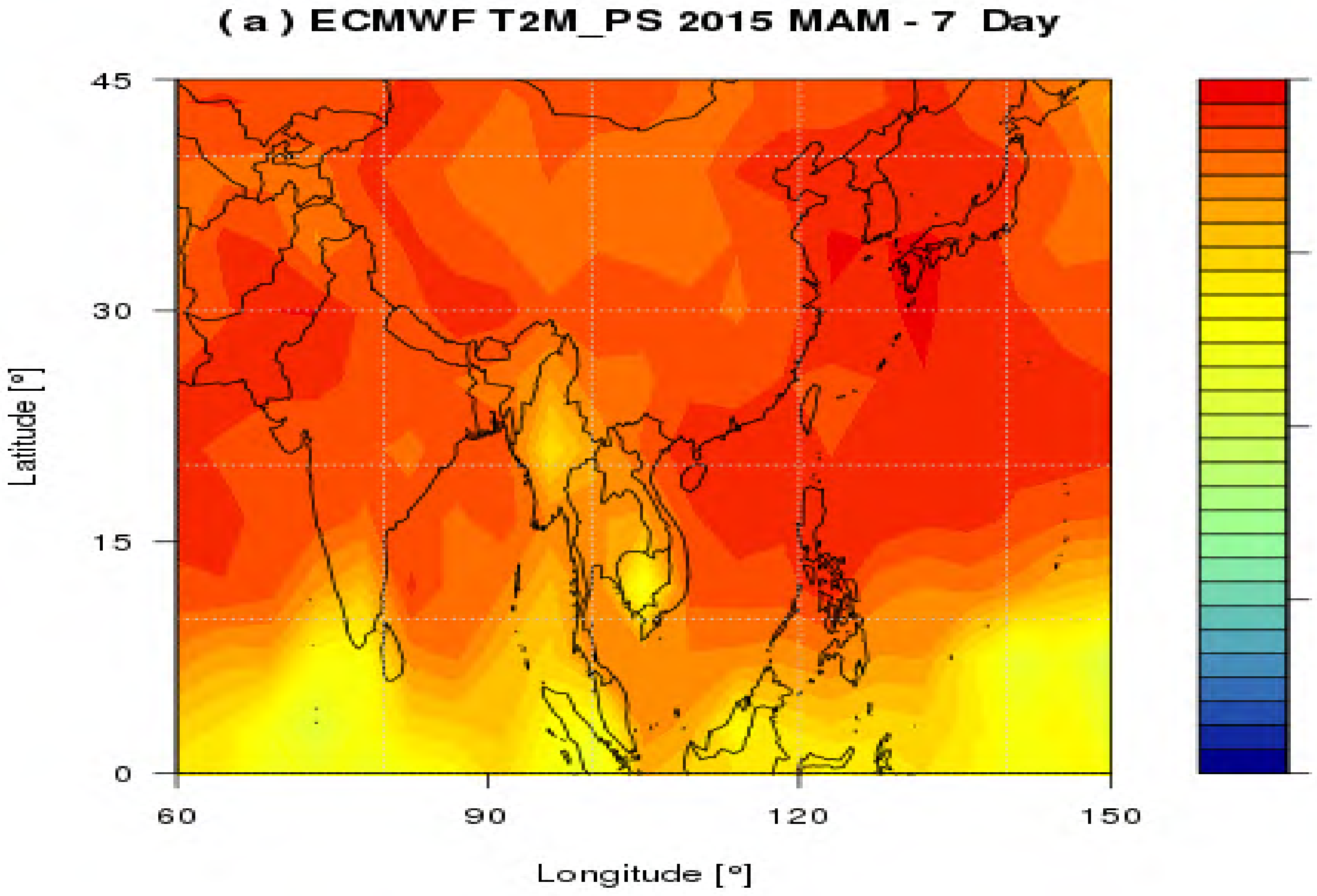
ECMWF S2S 2015 Spring TOTPRCP SS

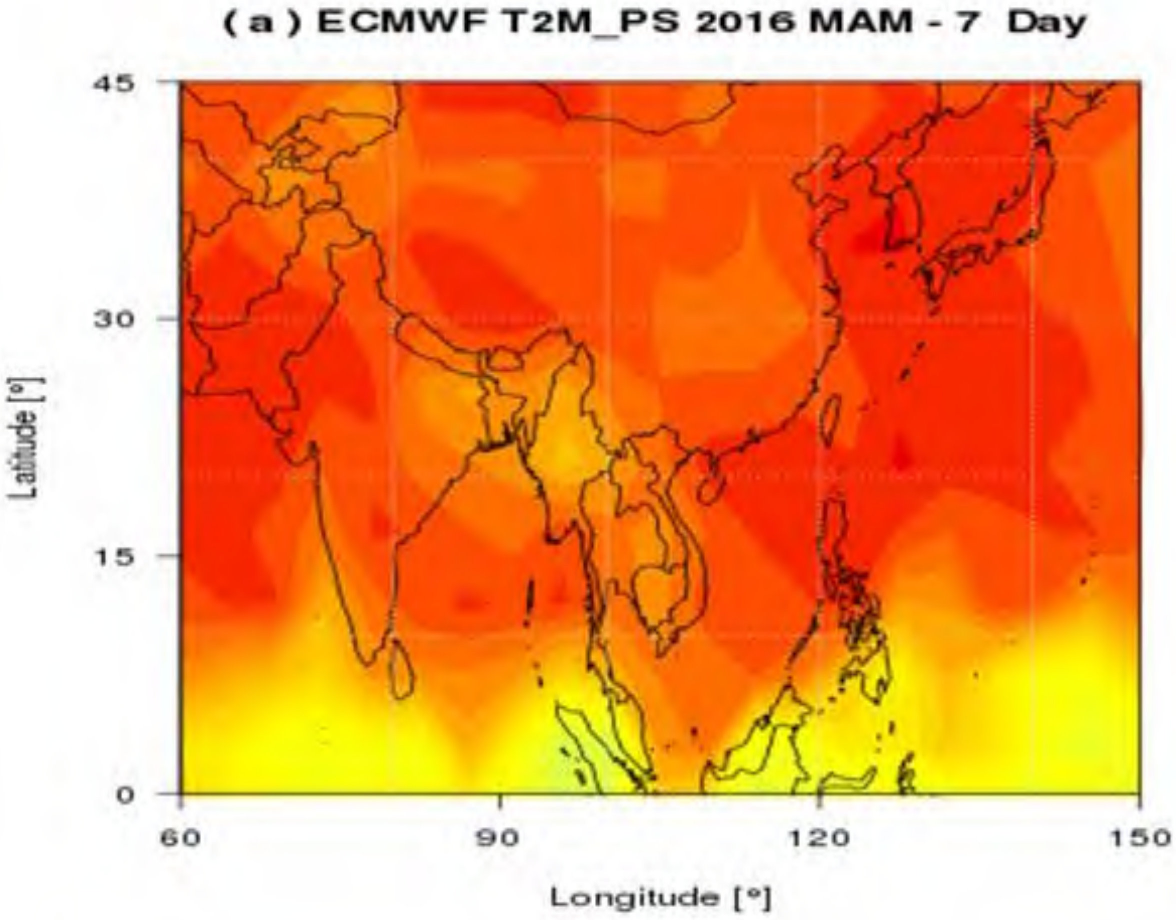


ECMWF S2S 2015 Spring TOTPRCP SS

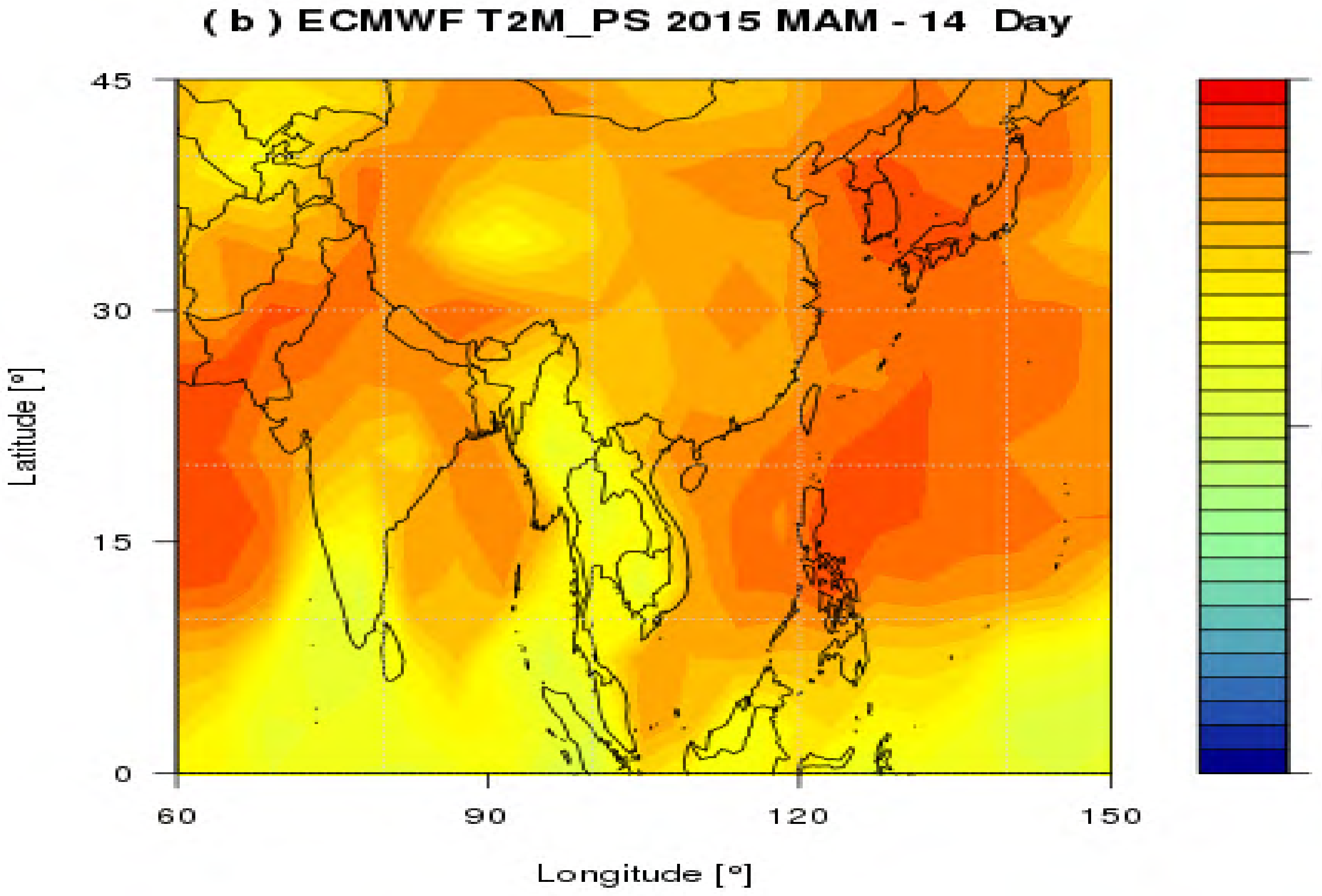


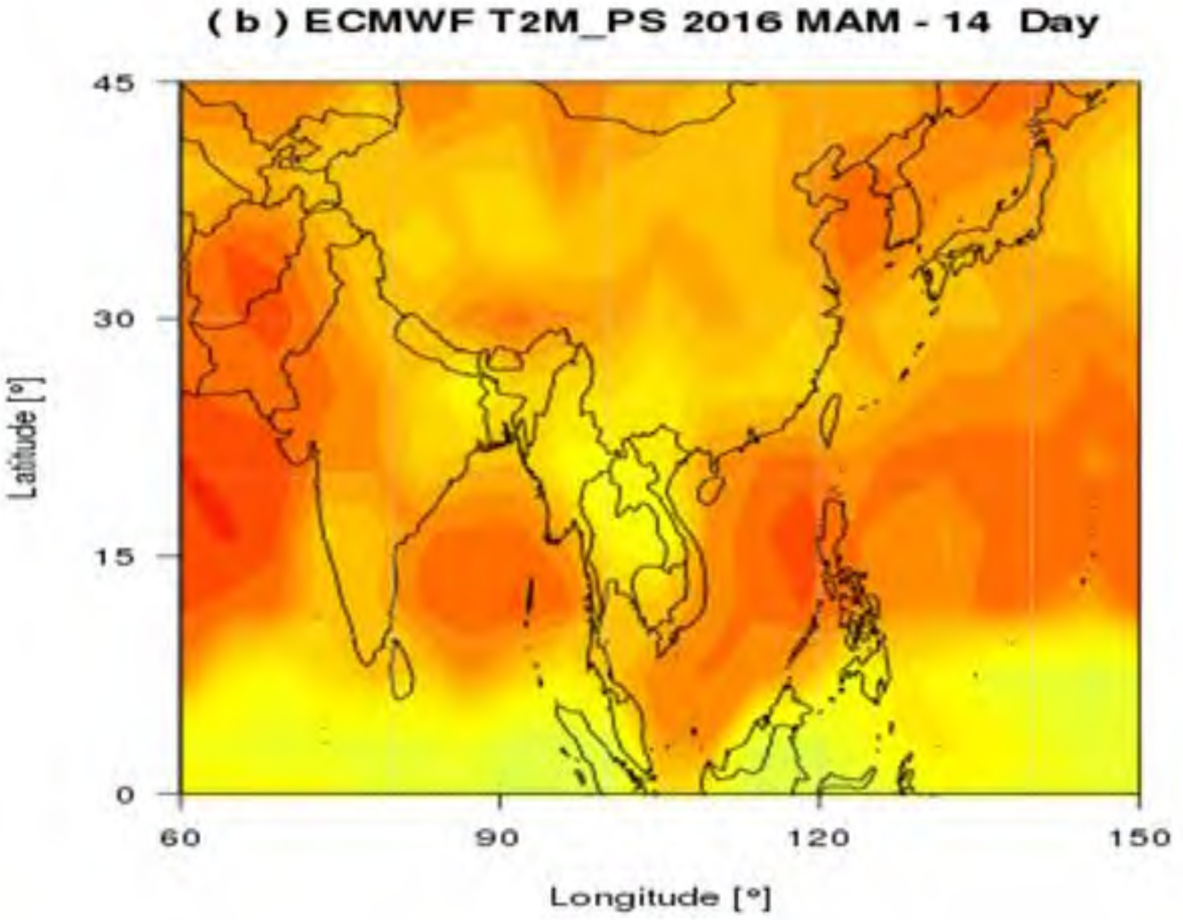
ECMWF S2S 2015 VS 2016 Spring T2M PS



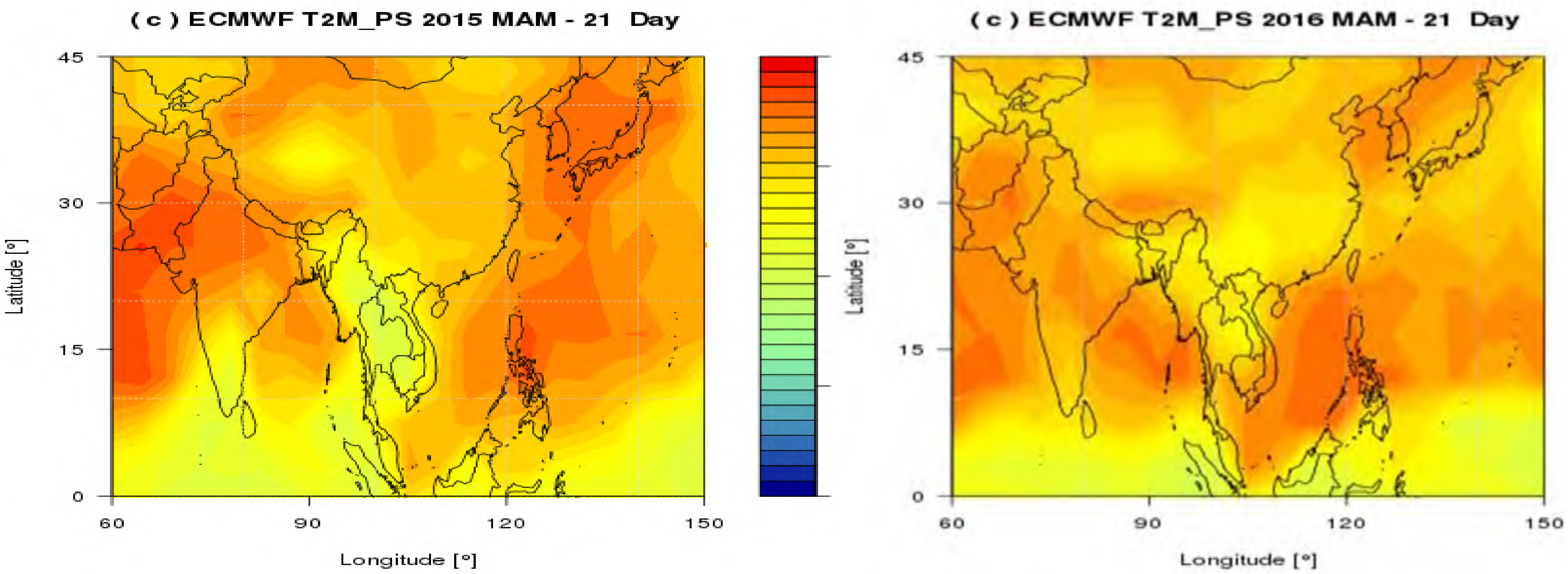


ECMWF S2S 2015 VS 2016 Spring T2M PS

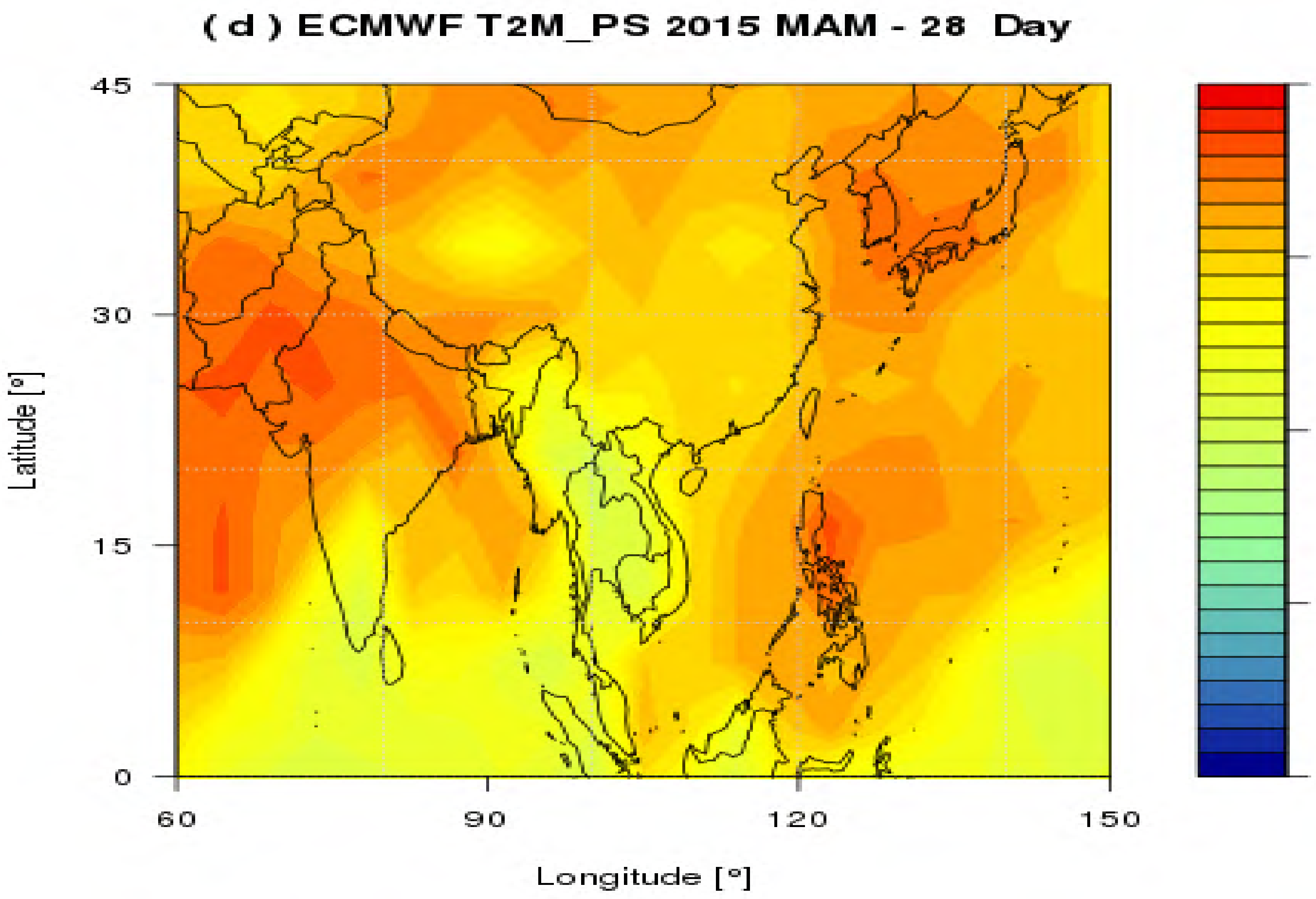


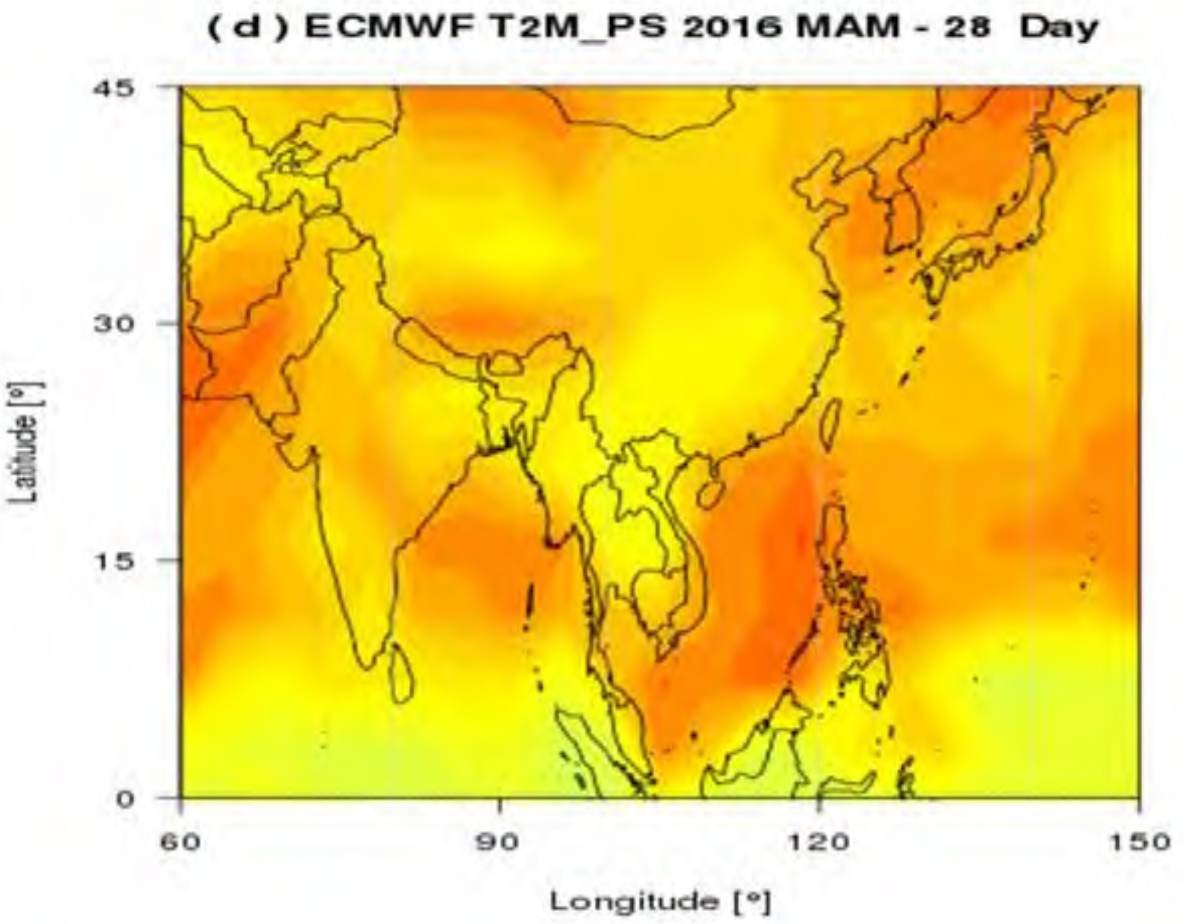


ECMWF S2S 2015 VS 2016 Spring T2M PS

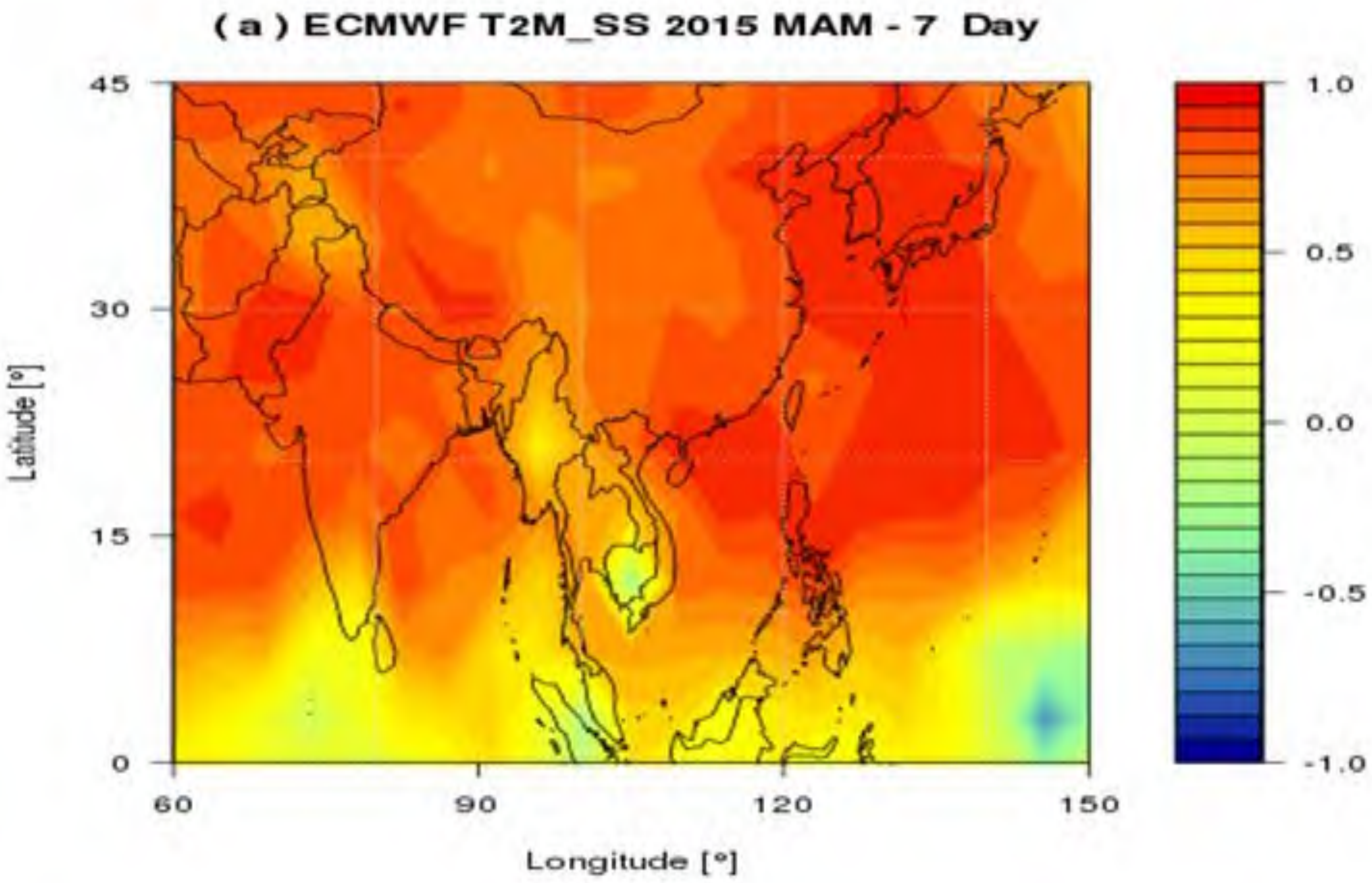


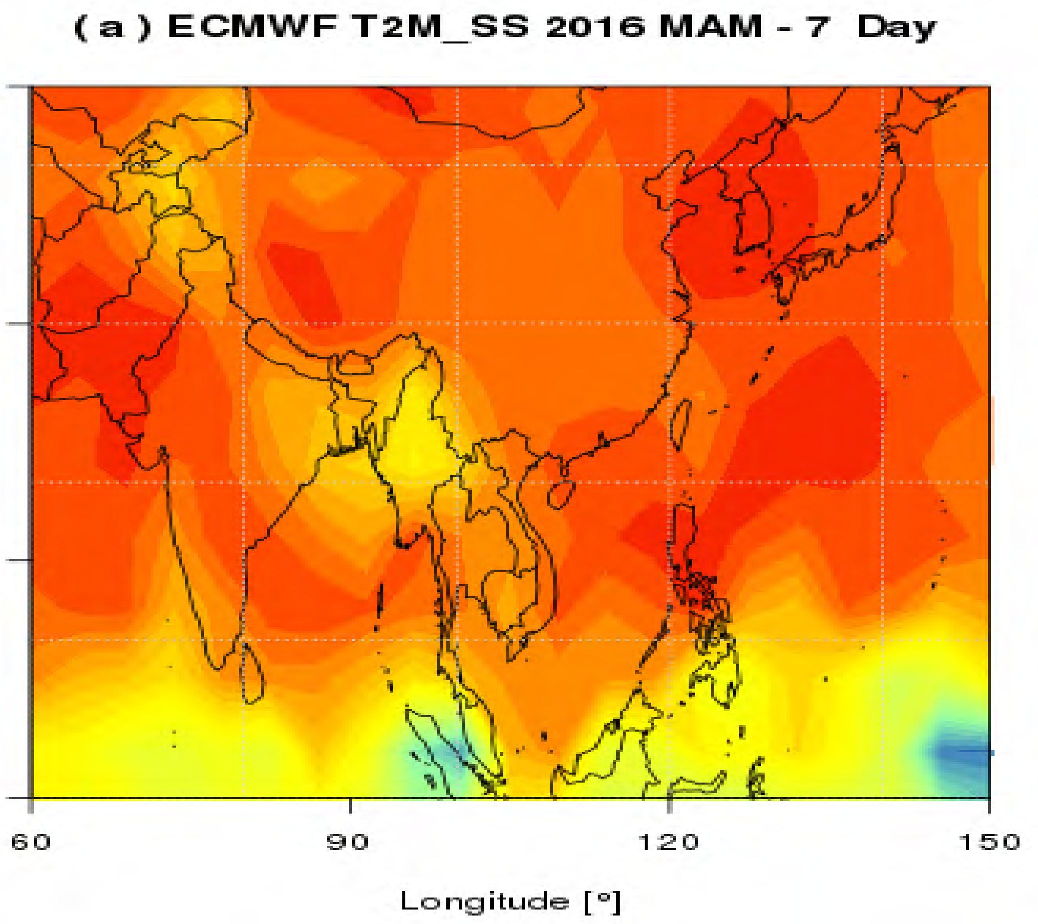
ECMWF S2S 2015 VS 2016 Spring T2M PS



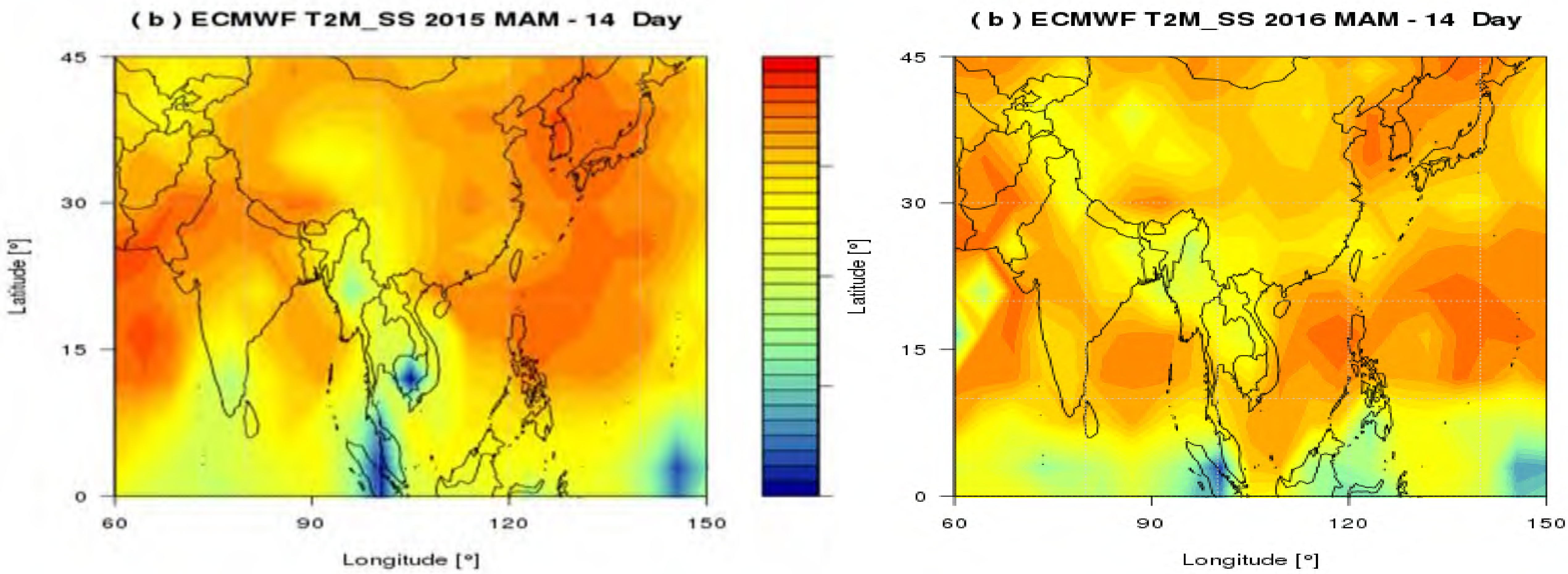


ECMWF S2S 2015 VS 2016 Spring T2M SS

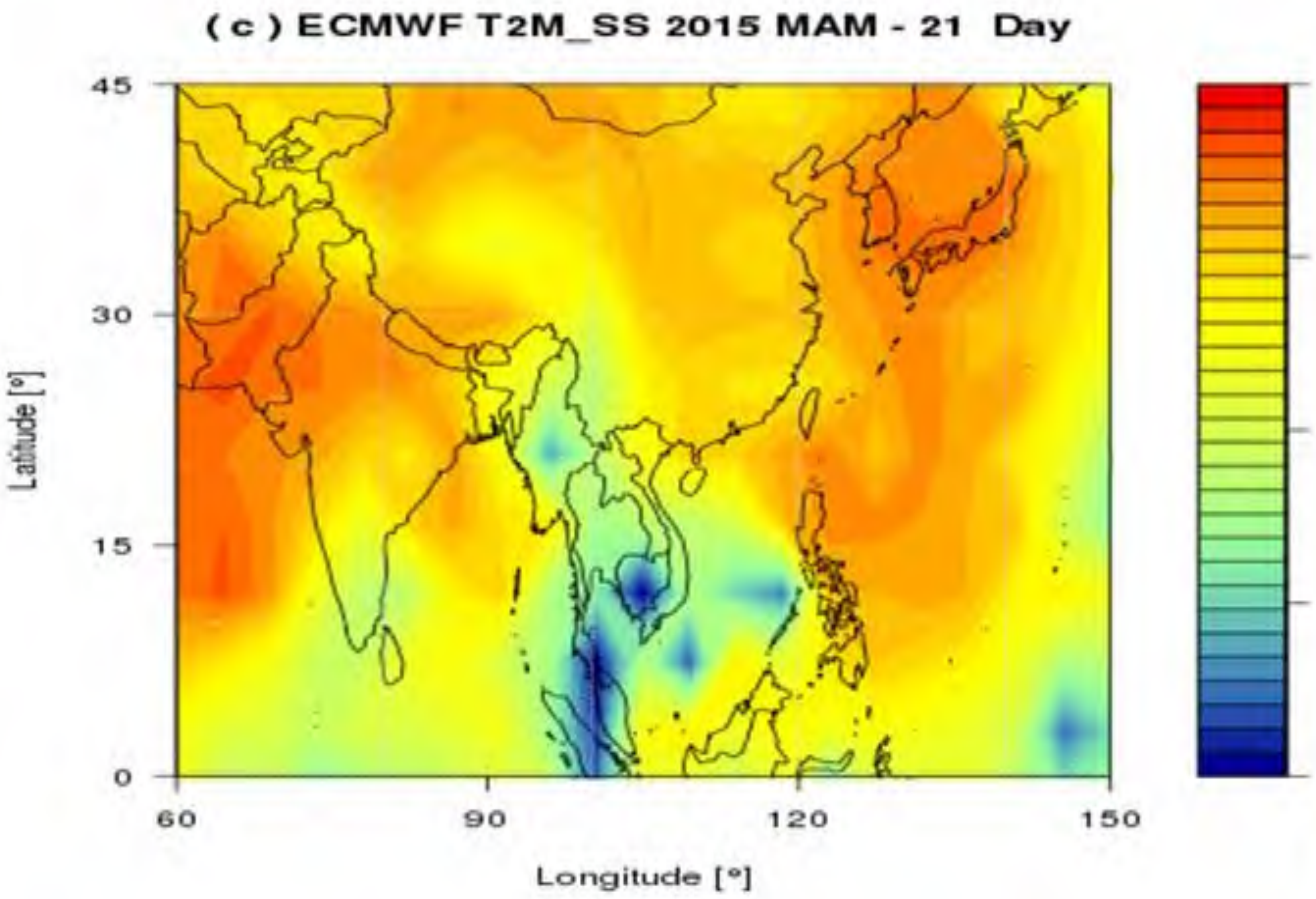


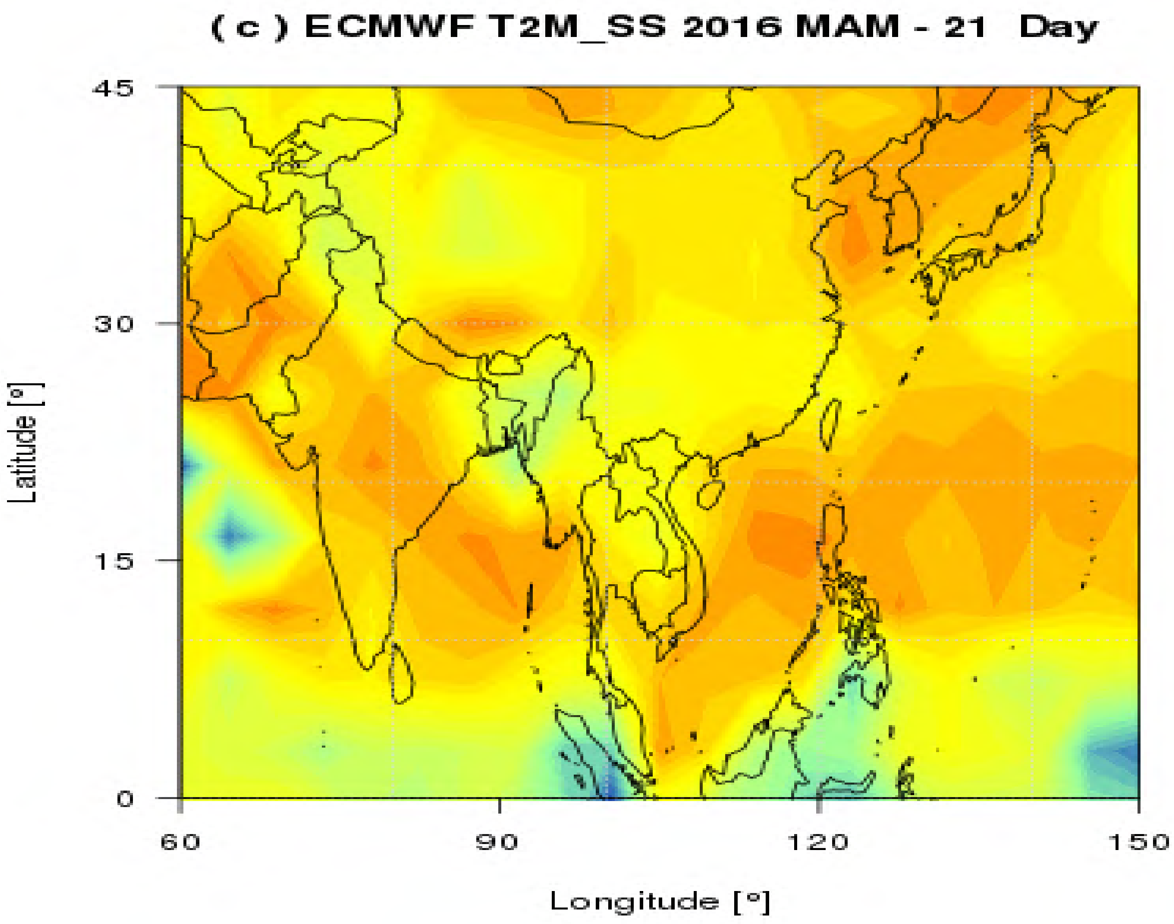


ECMWF S2S 2015 VS 2016 Spring T2M SS

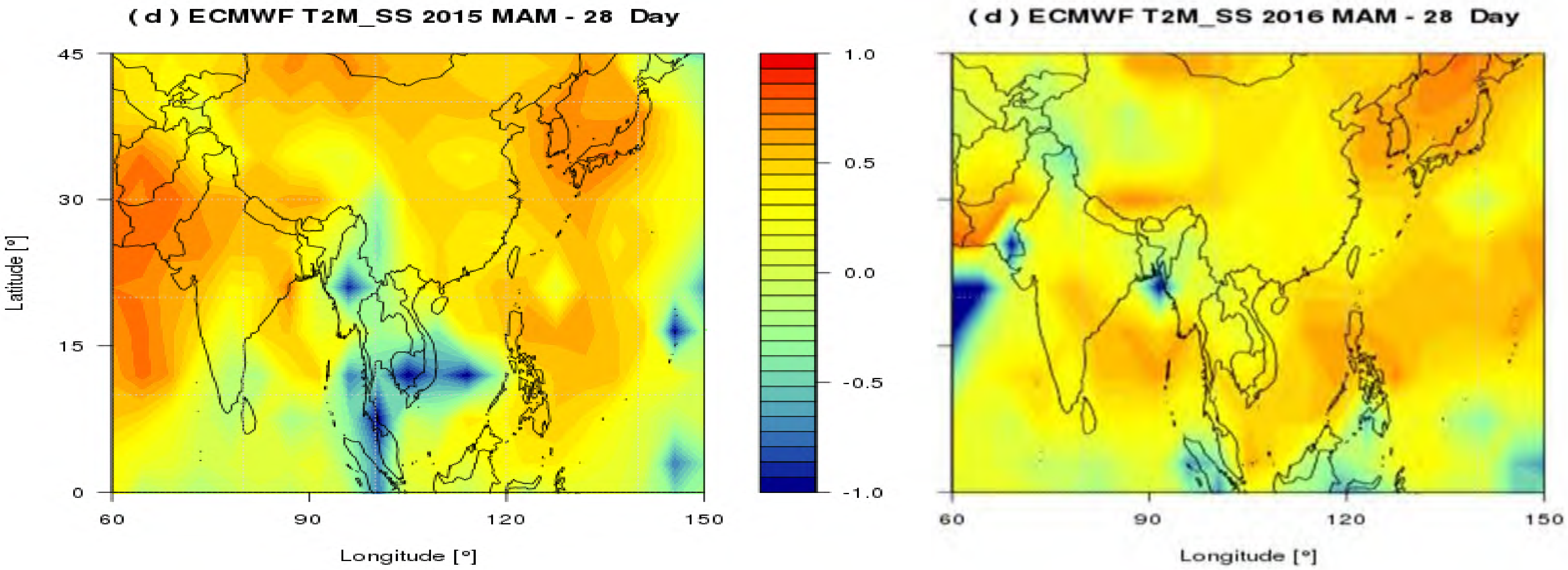


ECMWF S2S 2015 VS 2016 Spring T2M SS

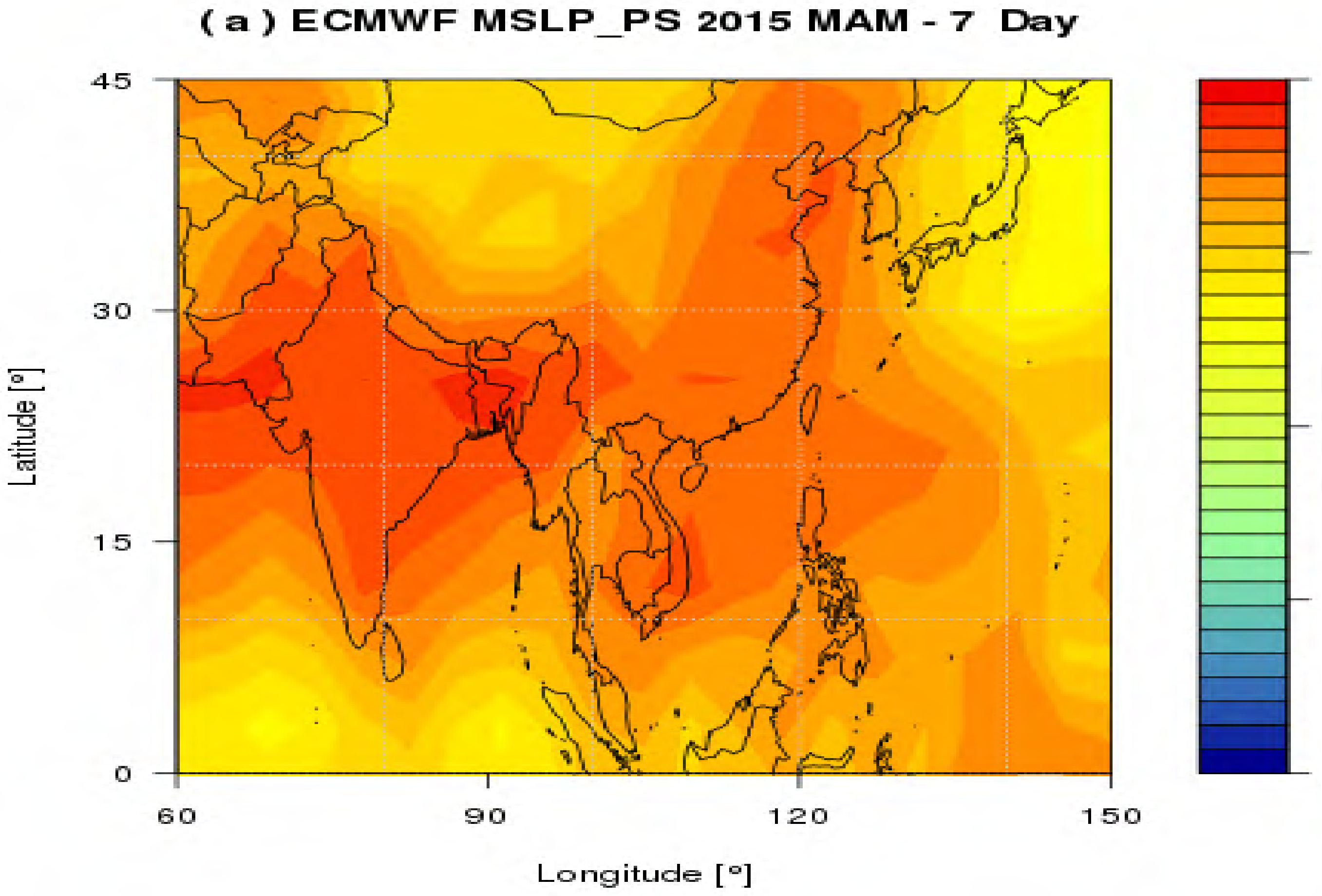


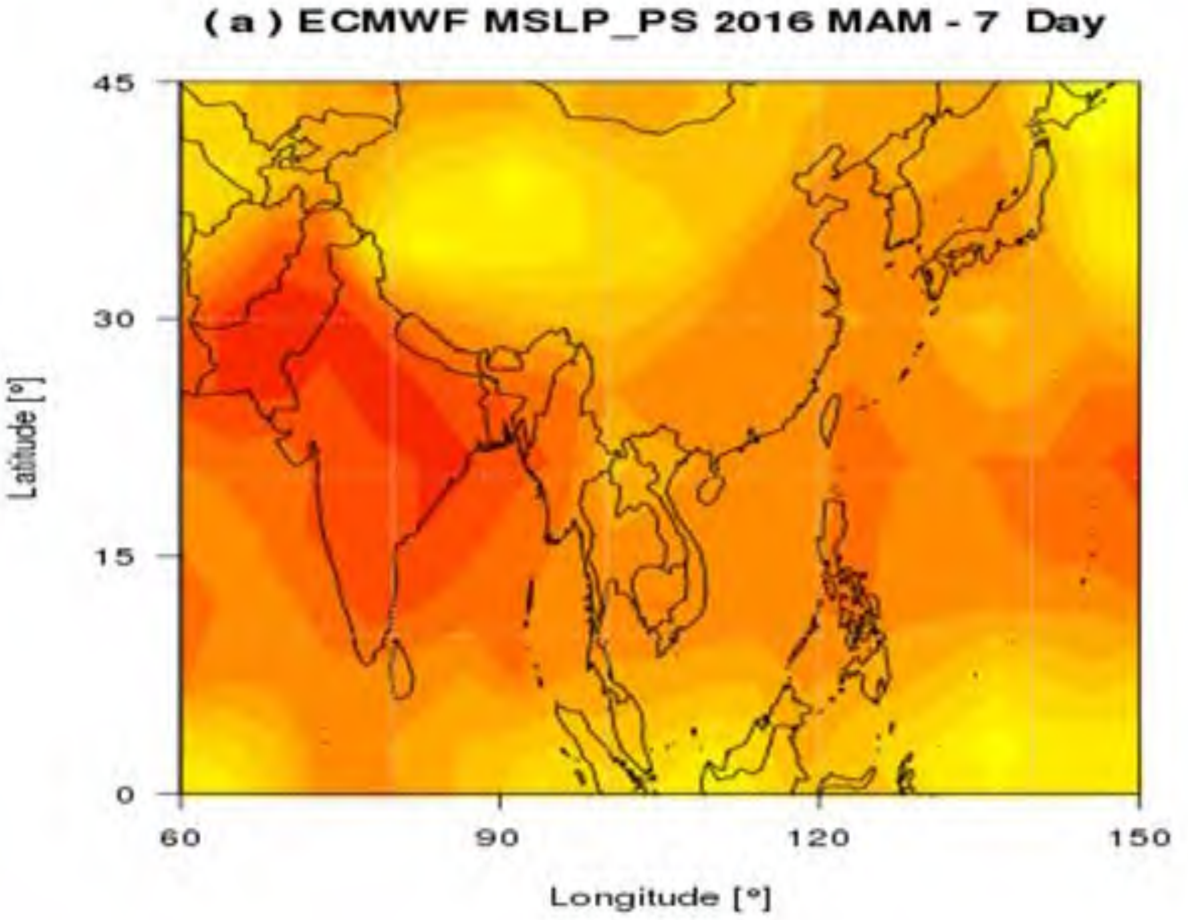


ECMWF S2S 2015 VS 2016 Spring T2M SS

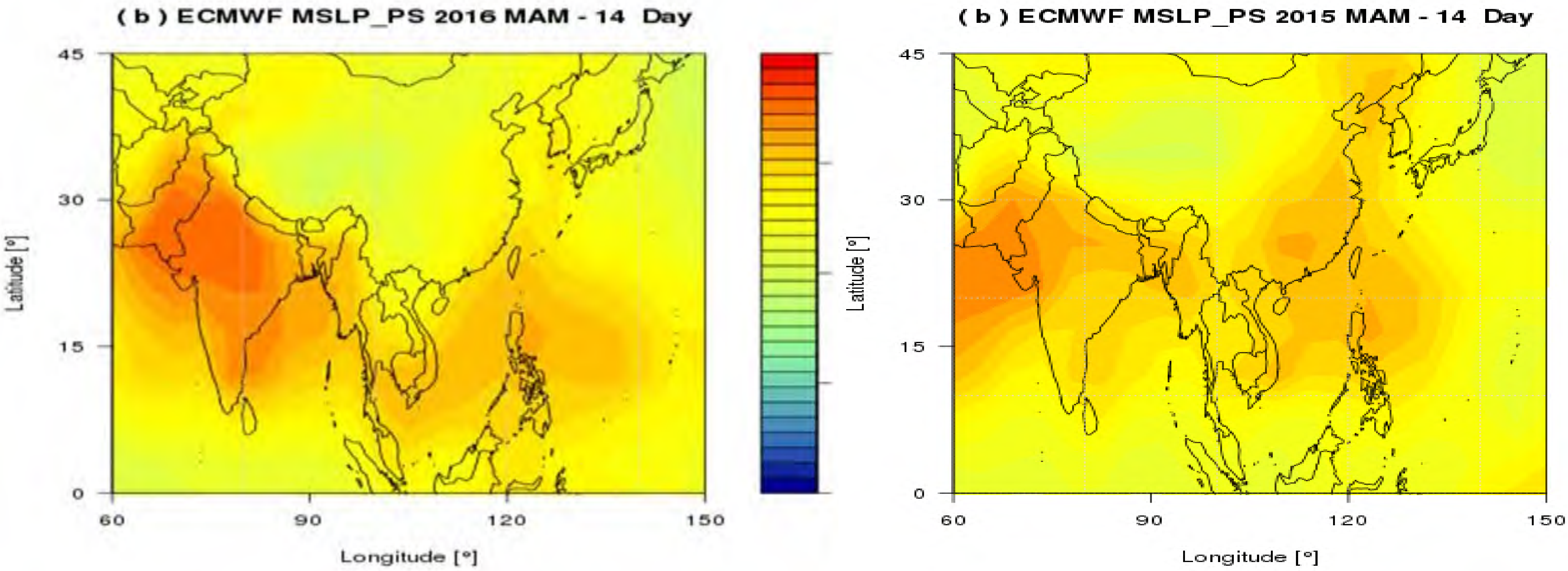


ECMWF S2S 2015 VS 2016 Spring MSLP PS

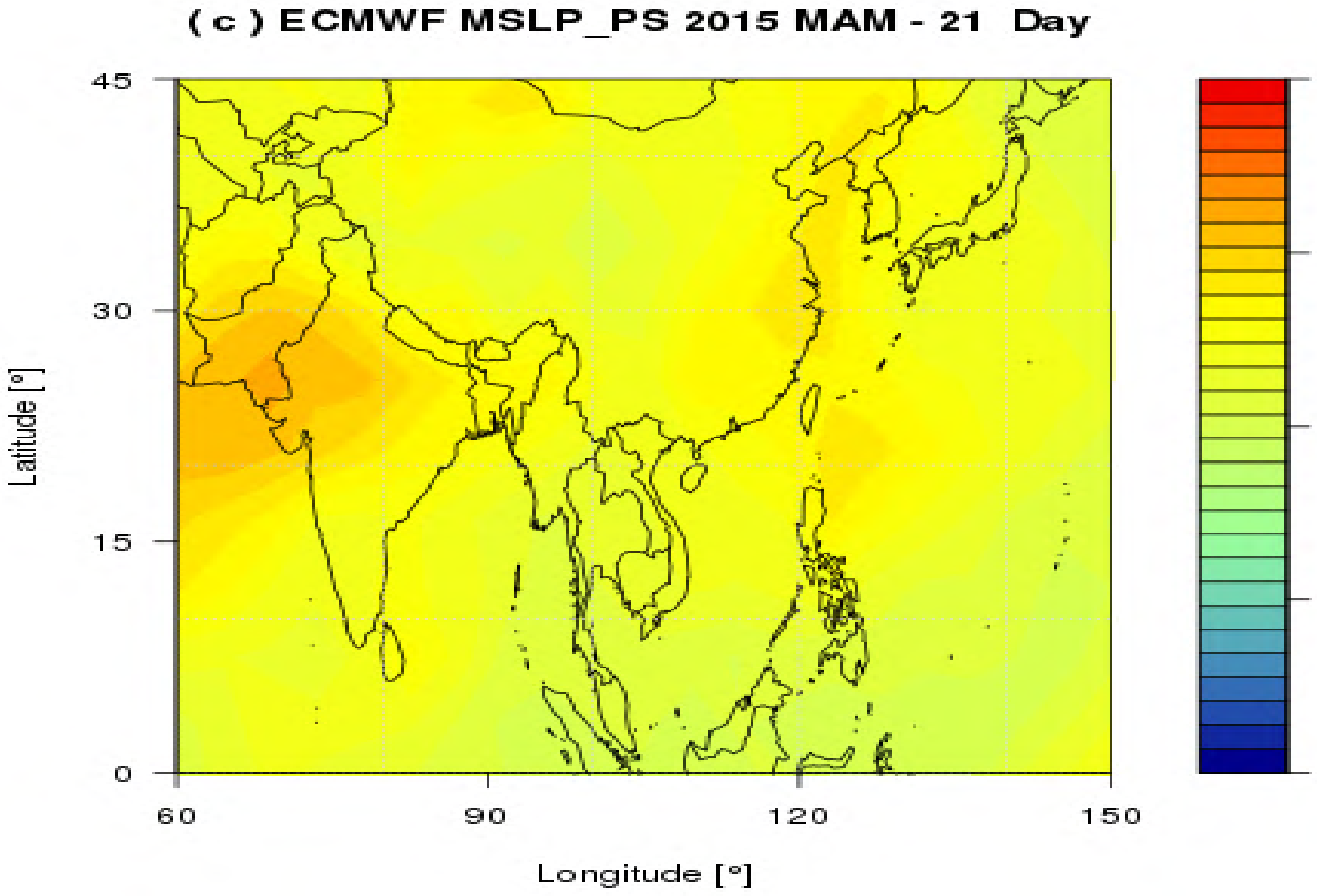


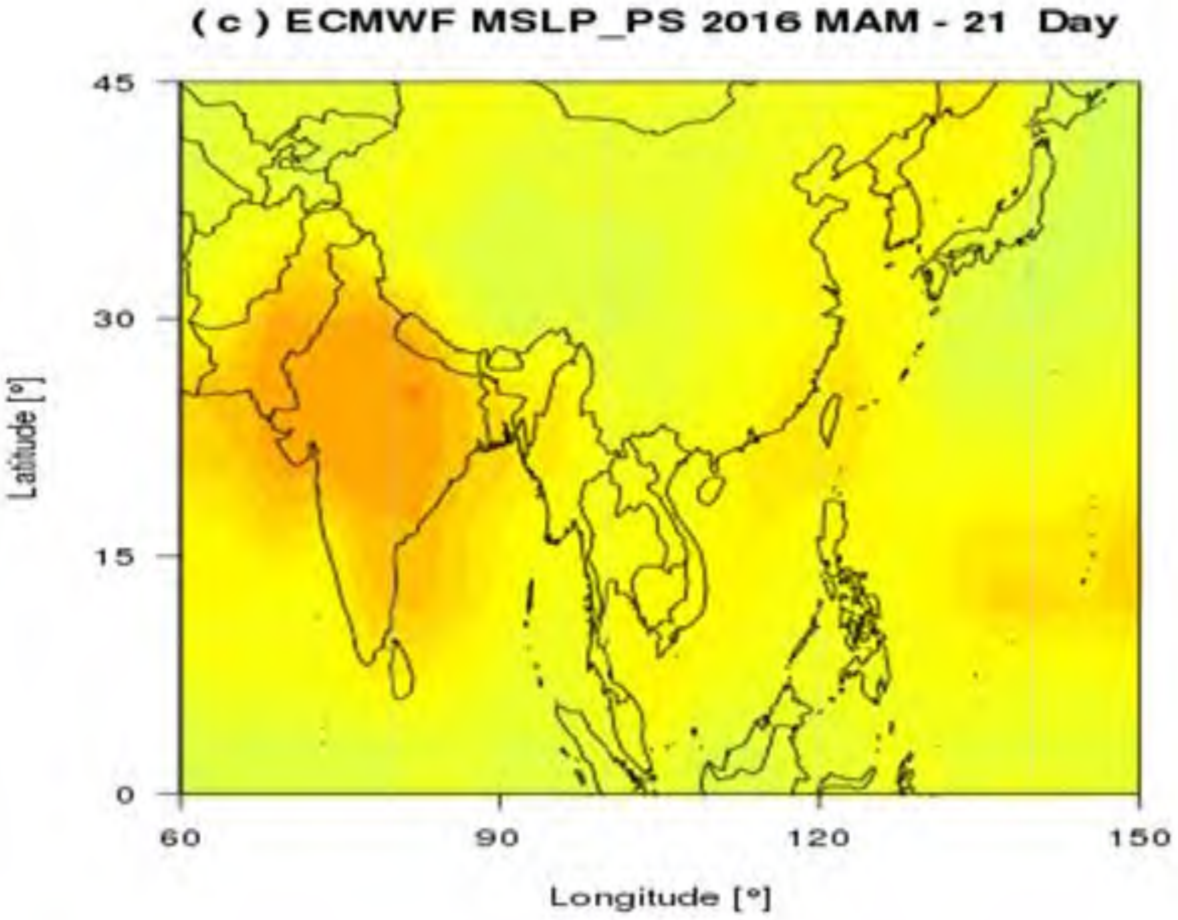


ECMWF S2S 2015 VS 2016 Spring MSLP PS

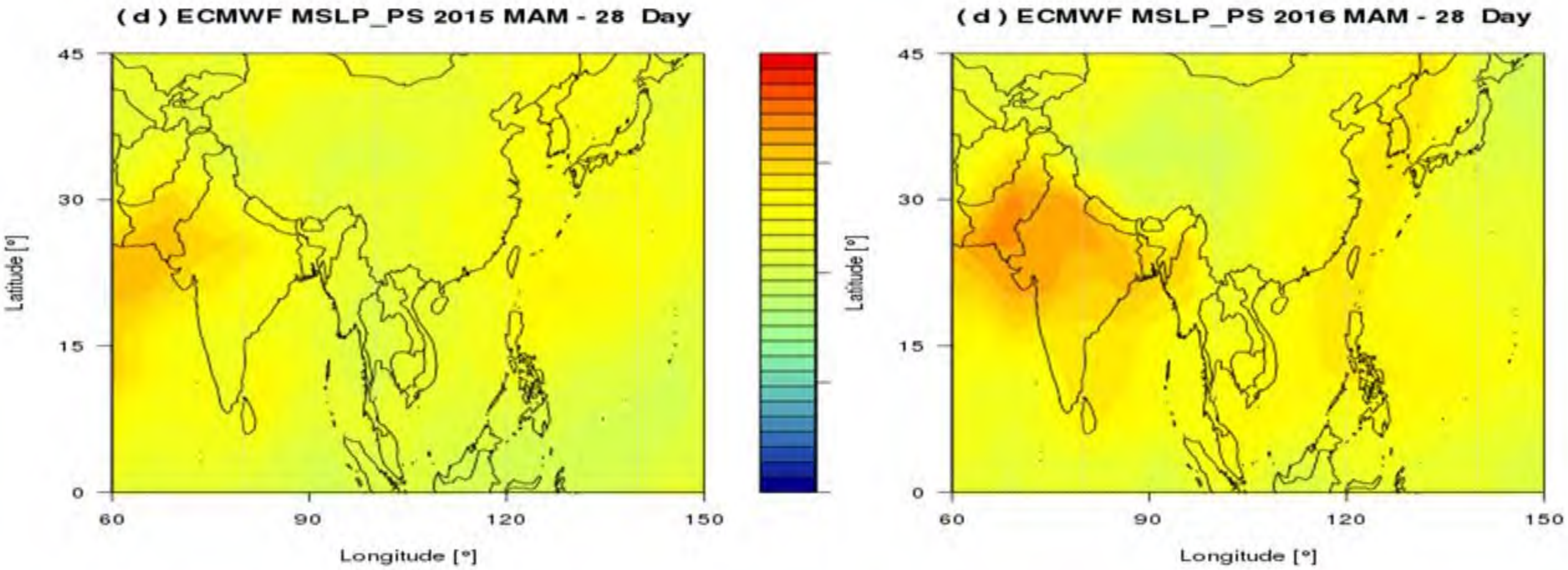


ECMWF S2S 2015 VS 2016 Spring MSLP PS

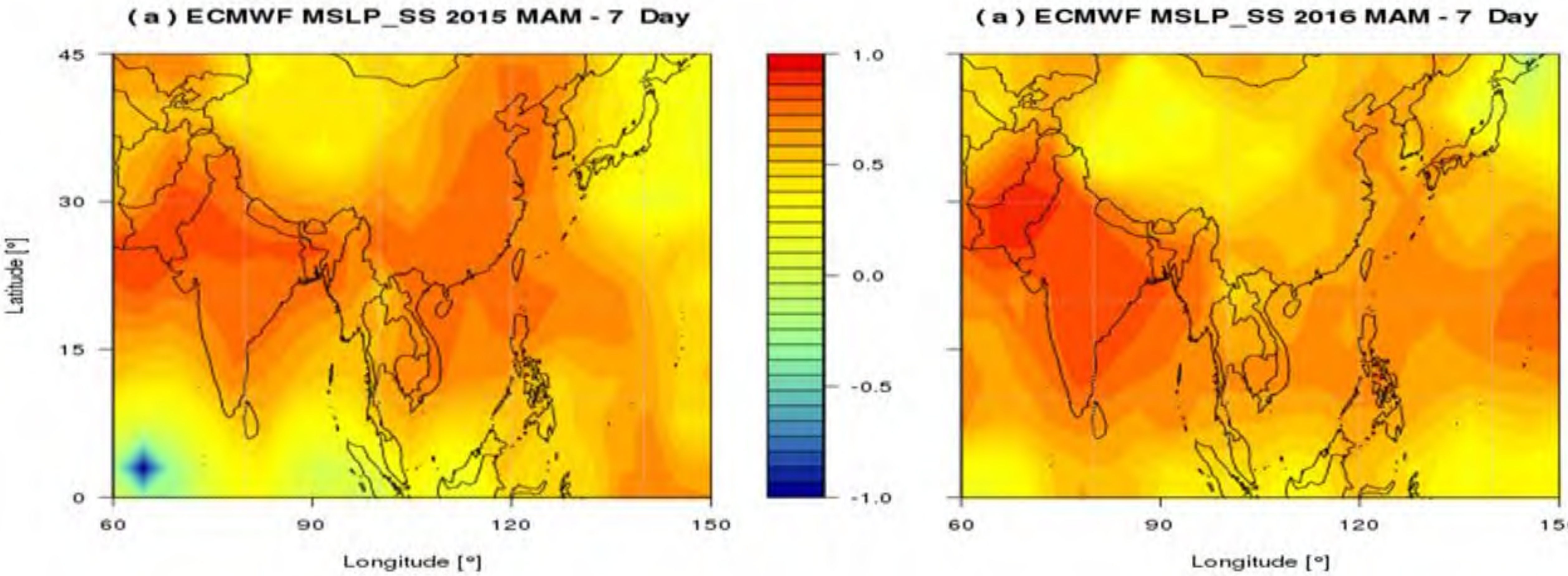




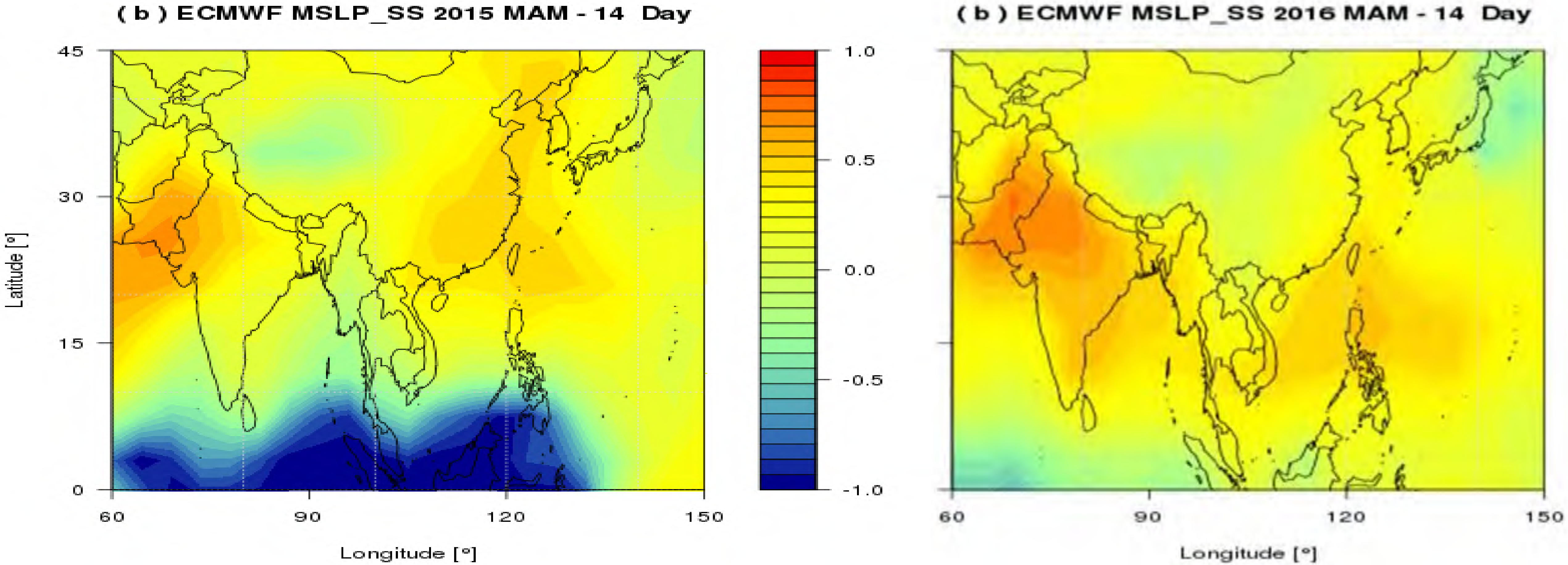
ECMWF S2S 2015 VS 2016 Spring MSLP PS



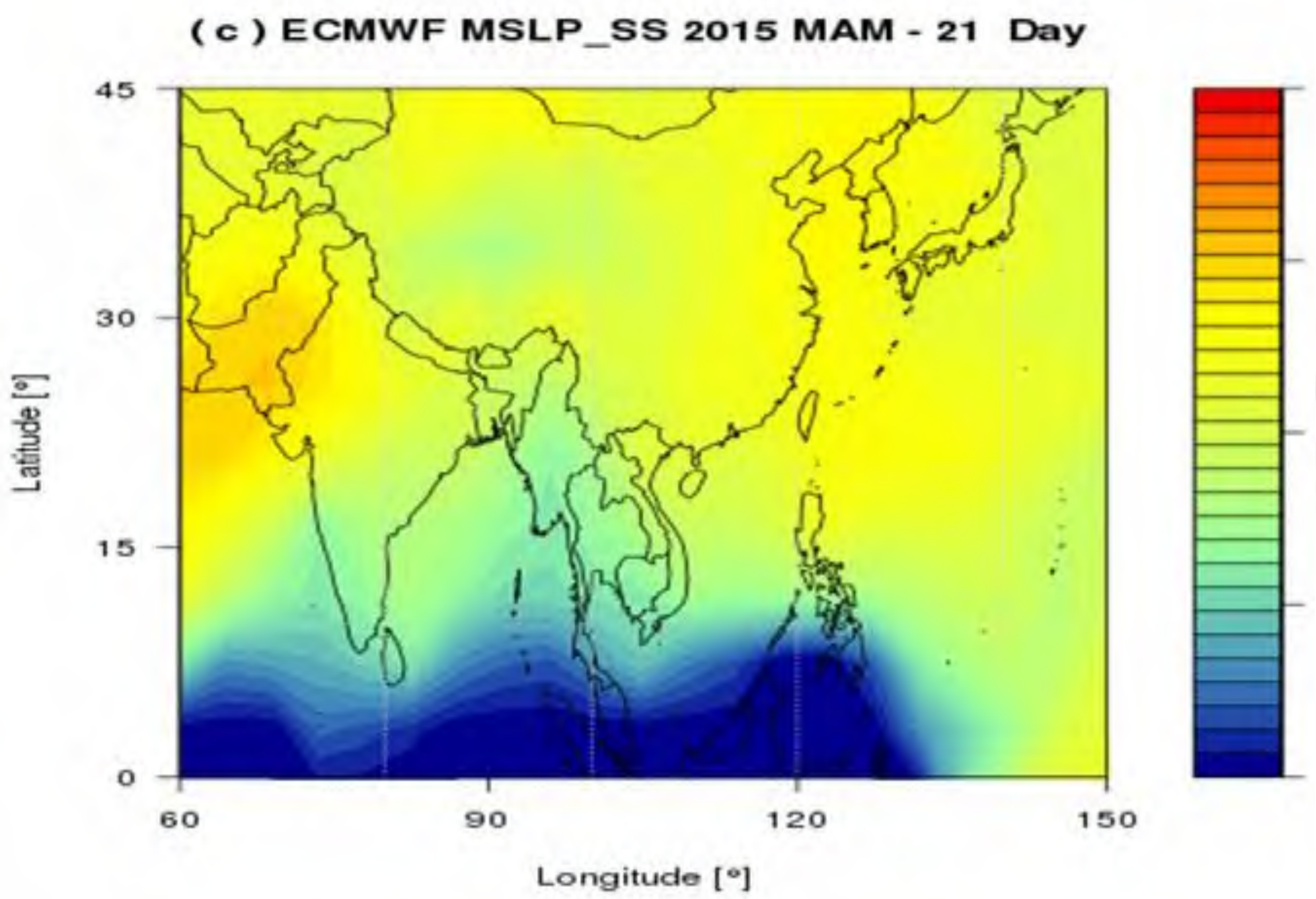
ECMWF S2S 2015 VS 2016 Spring MSLP SS

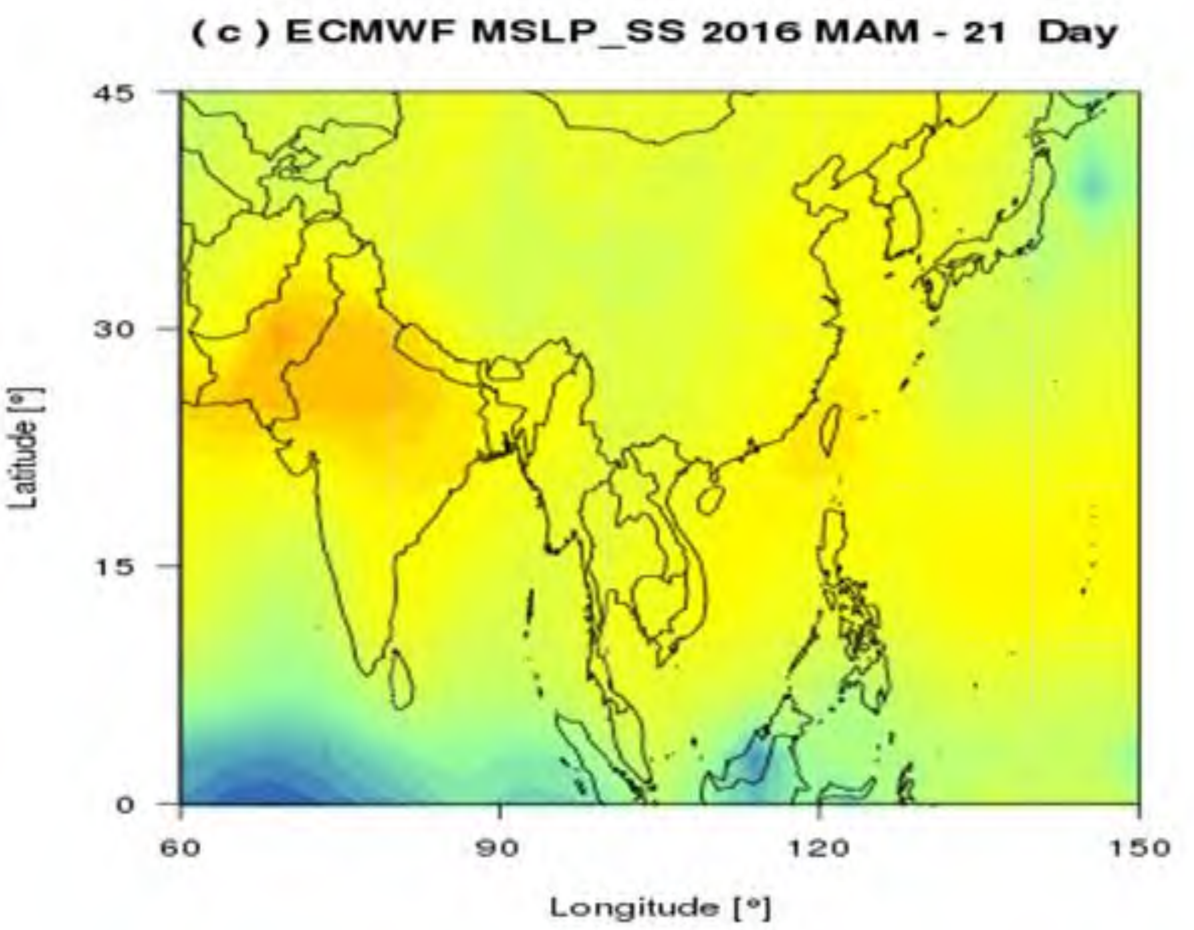


ECMWF S2S 2015 VS 2016 Spring MSLP SS

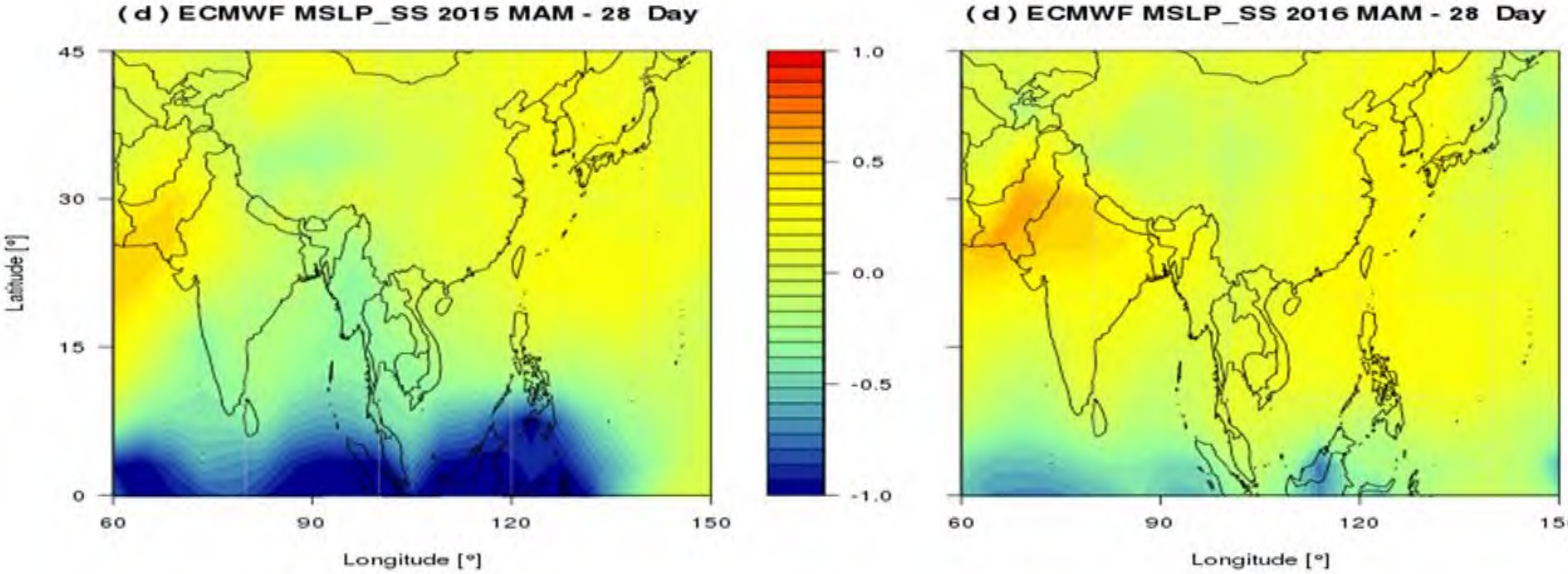


ECMWF S2S 2015 VS 2016 Spring MSLP SS

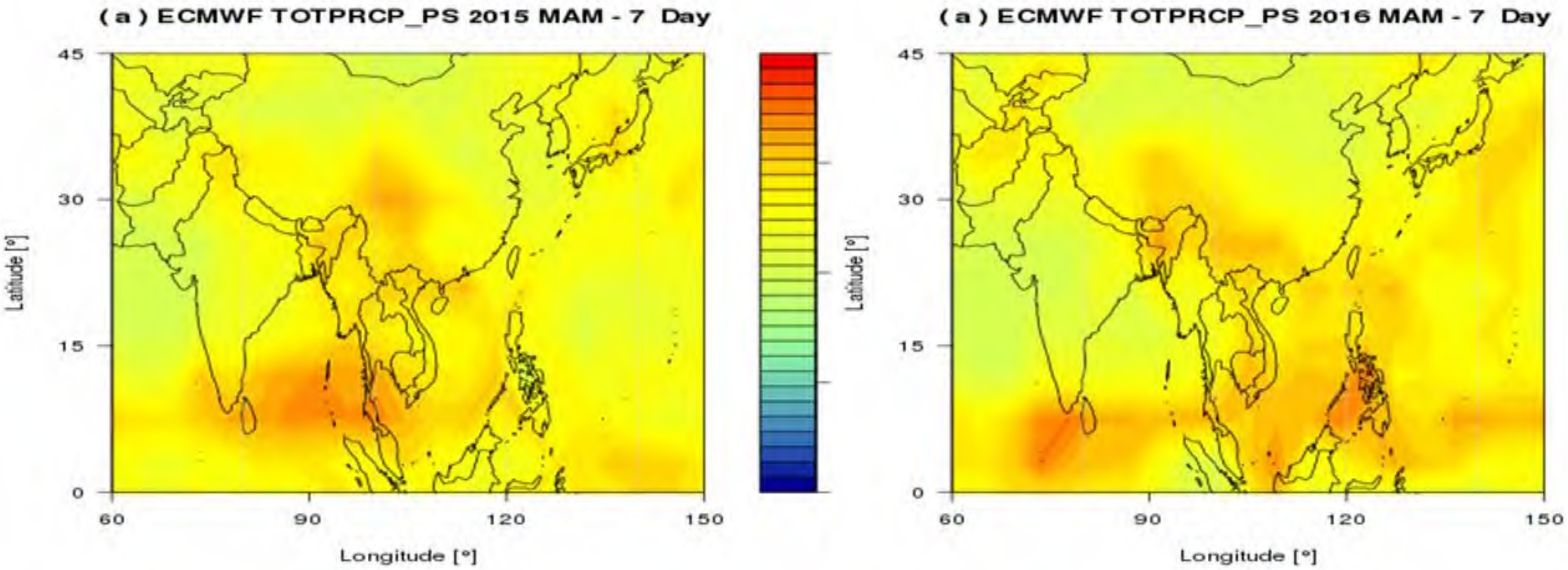




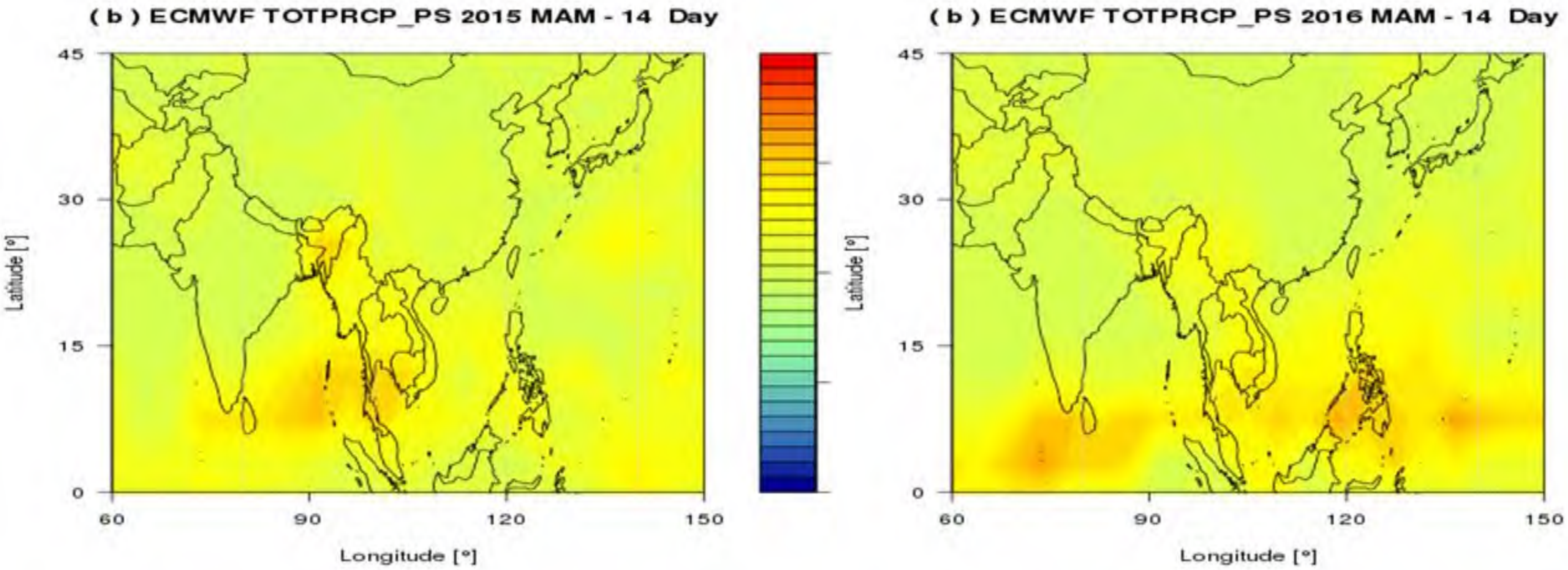
ECMWF S2S 2015 VS 2016 Spring MSLP SS



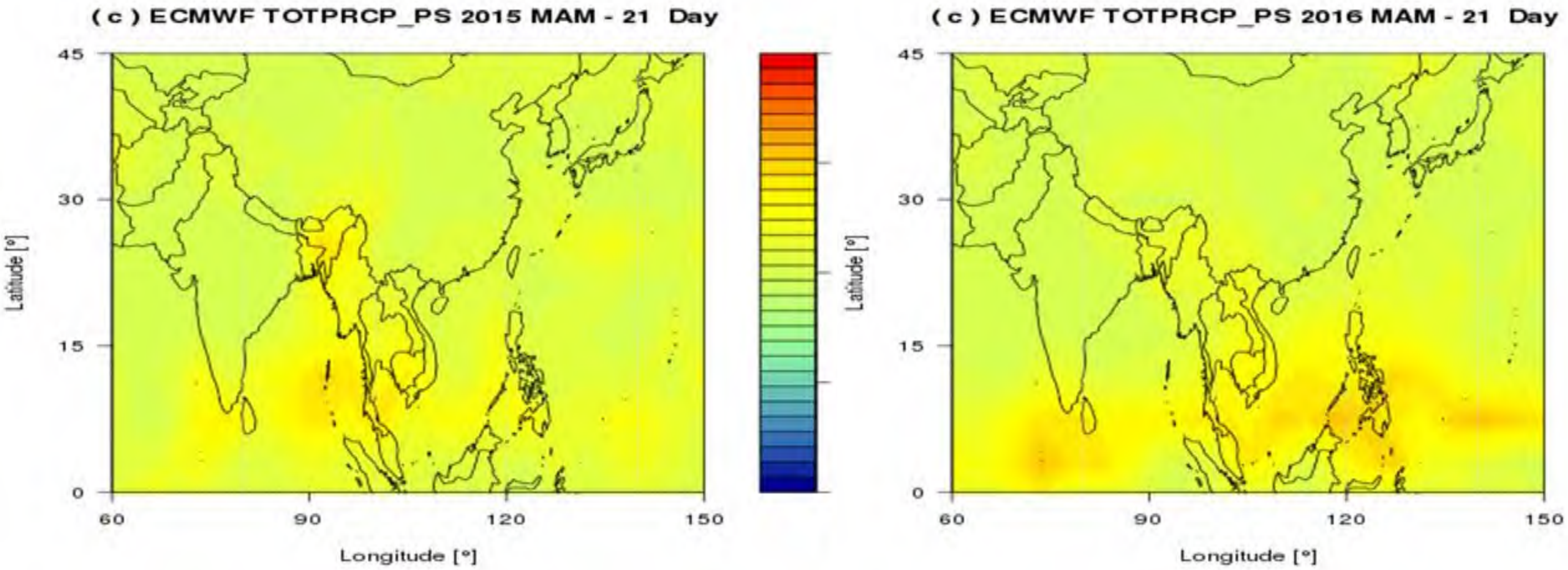
ECMWF S2S 2015 VS 2016 Spring TOTPRCP PS



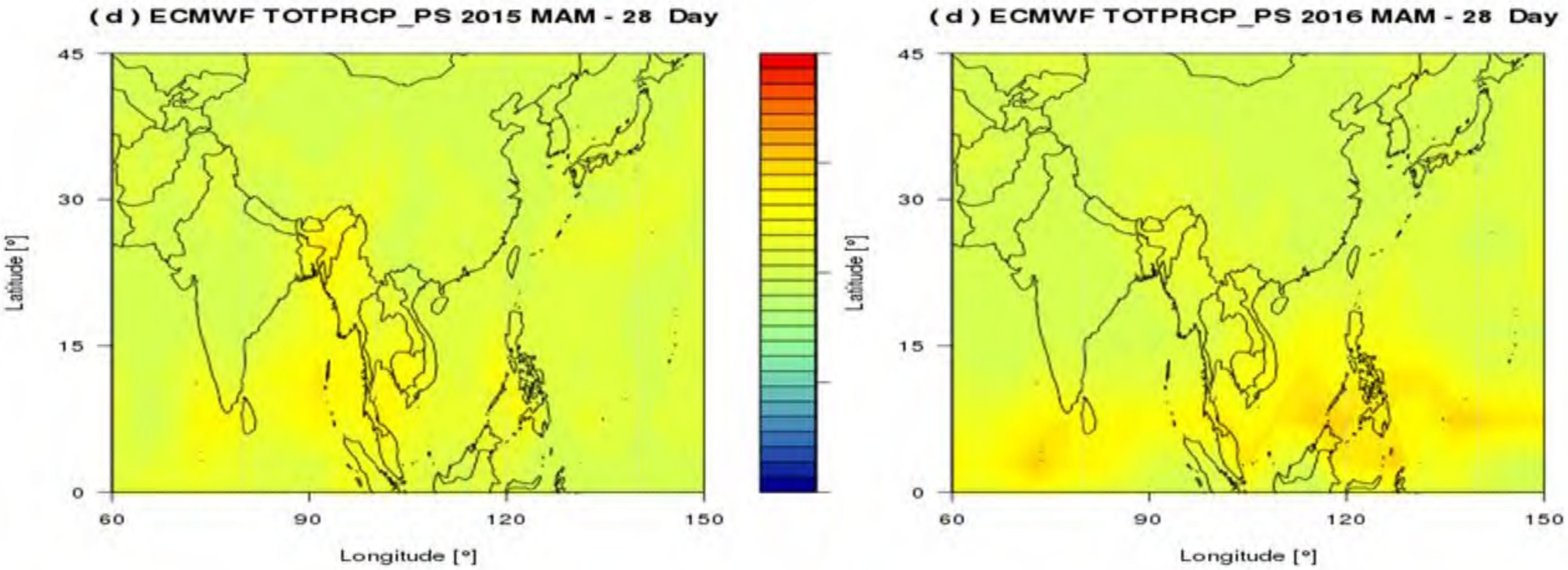
ECMWF S2S 2015 VS 2016 Spring TOTPRCP PS



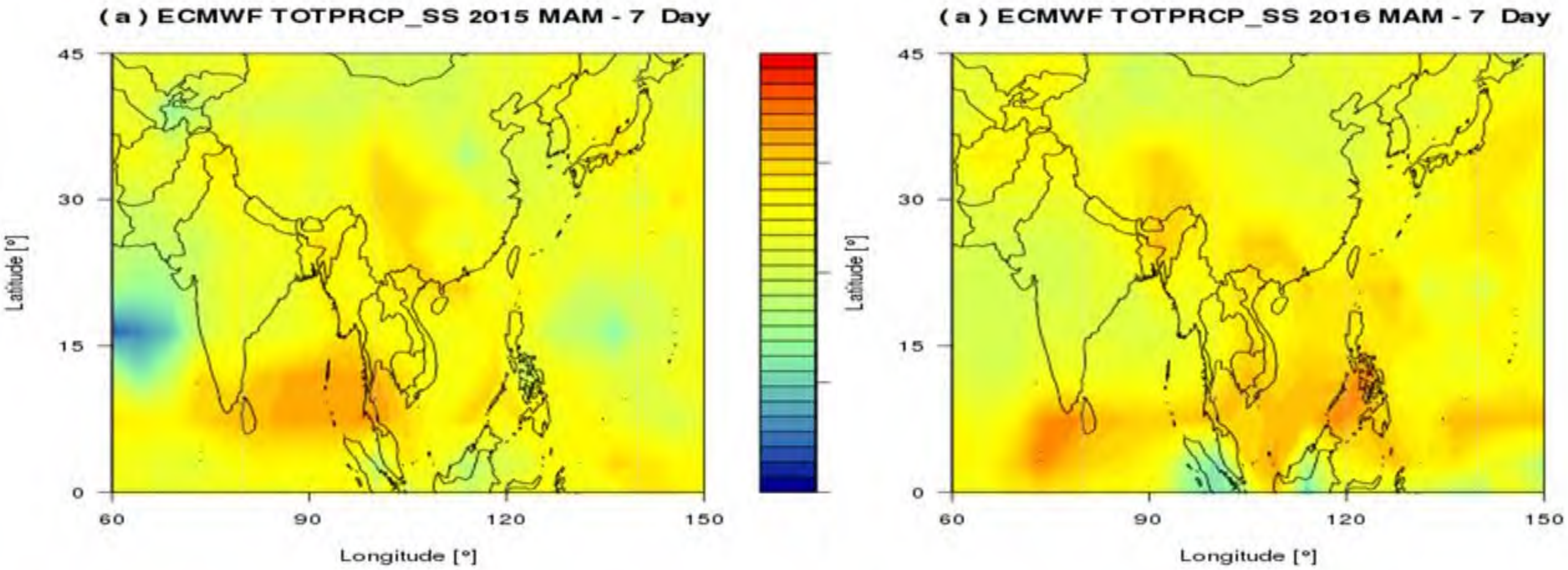
ECMWF S2S 2015 VS 2016 Spring TOTPRCP PS



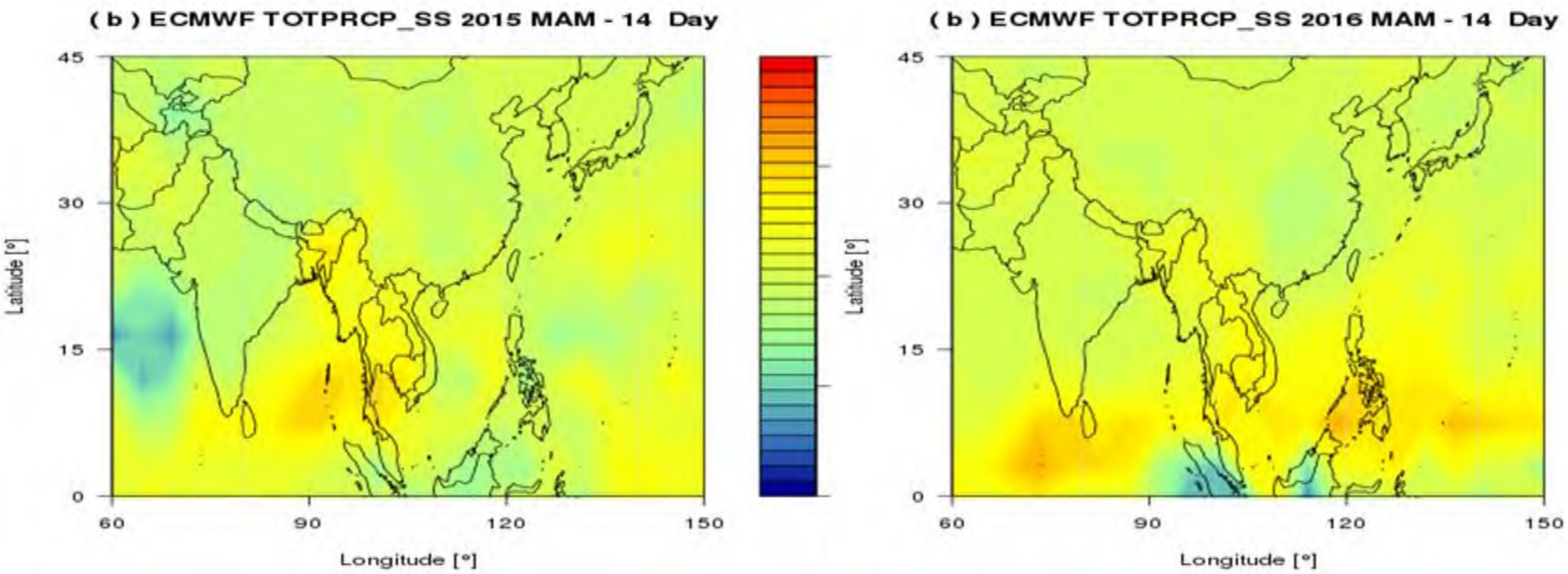
ECMWF S2S 2015 VS 2016 Spring TOTPRCP PS



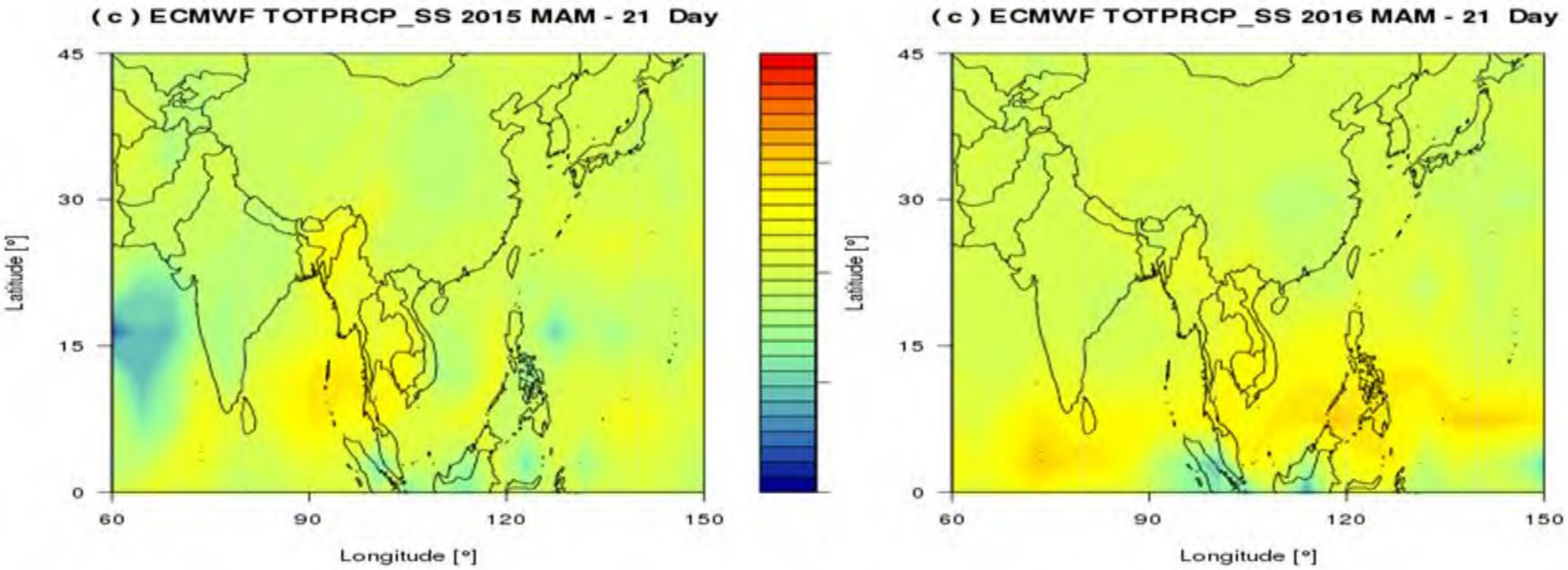
ECMWF S2S 2015 VS 2016 Spring TOTPRCP SS



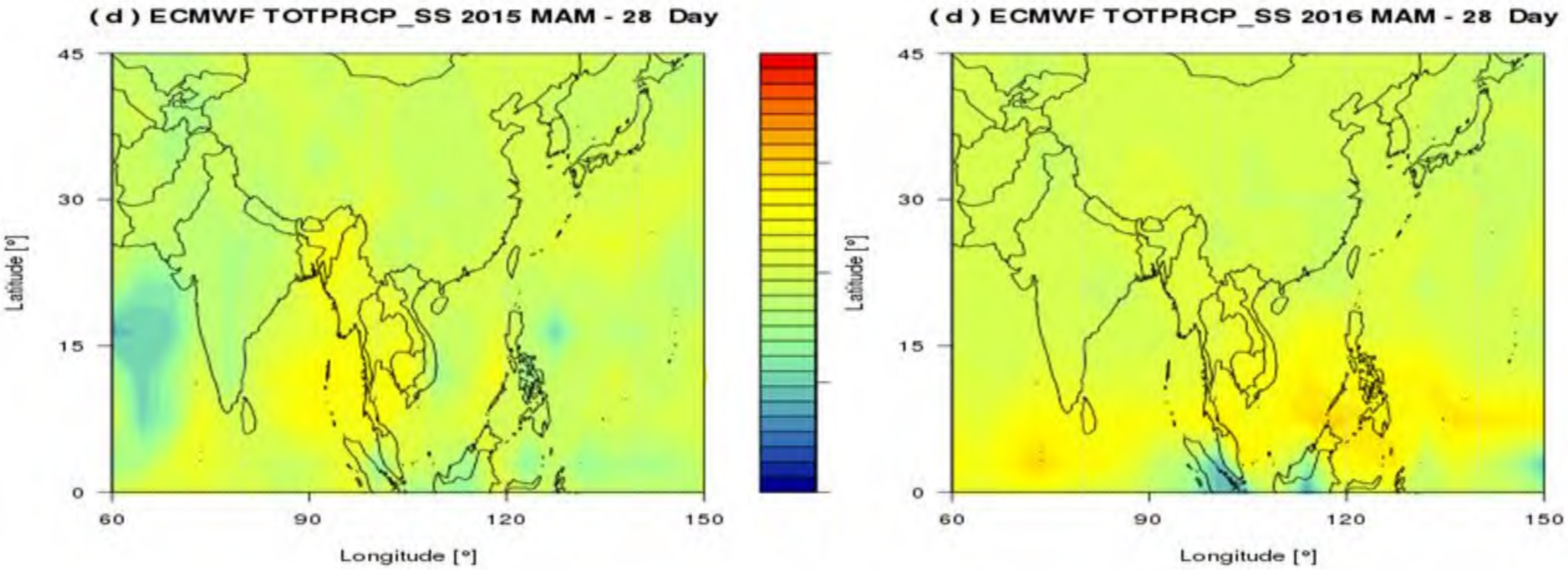
ECMWF S2S 2015 VS 2016 Spring TOTPRCP SS



ECMWF S2S 2015 VS 2016 Spring TOTPRCP SS



ECMWF S2S 2015 VS 2016 Spring TOTPRCP SS



Summary

- In general, skill in winter is higher than it in summer.
- T2M Skill > MSLP Skill > TOTPRCP Skill
- In 2015 Spring, skills maintained better in Arabian Sea and India.
- In 2016 Spring, skills maintained better in western Pacific Ocean and South China Sea.
- ECMWF had better MSLP skill performance than NCEP in 2016 Spring.

Thank you for listening.

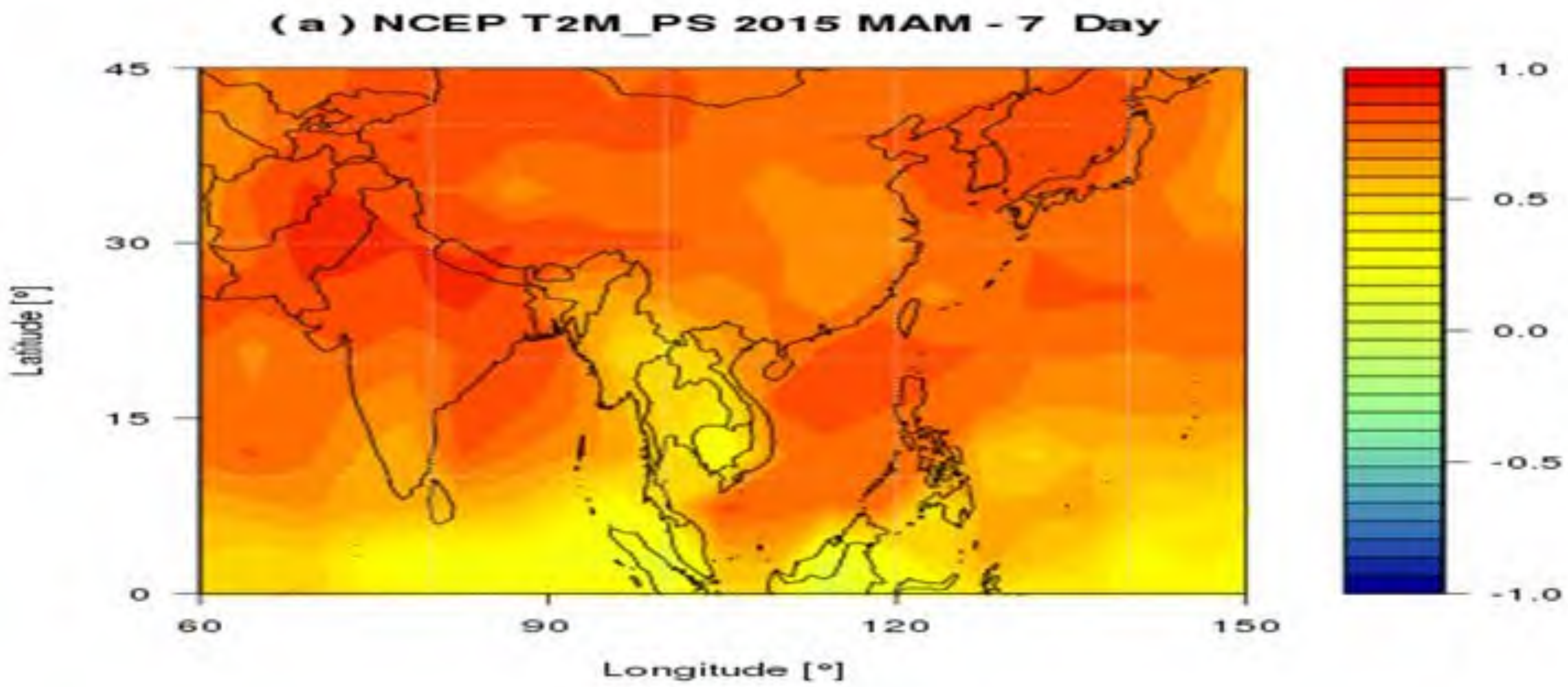
References

- Murphy AH, Epstein ES. 1989. Skill scores and correlation coefficients in model verification. Monthly Weather Review 117: 572–581.
- Bradley AA, Schwartz SS (2011) Summary verification measures and their interpretation for ensemble forecasts. Mon Weather Rev 139(9):3075–3089.

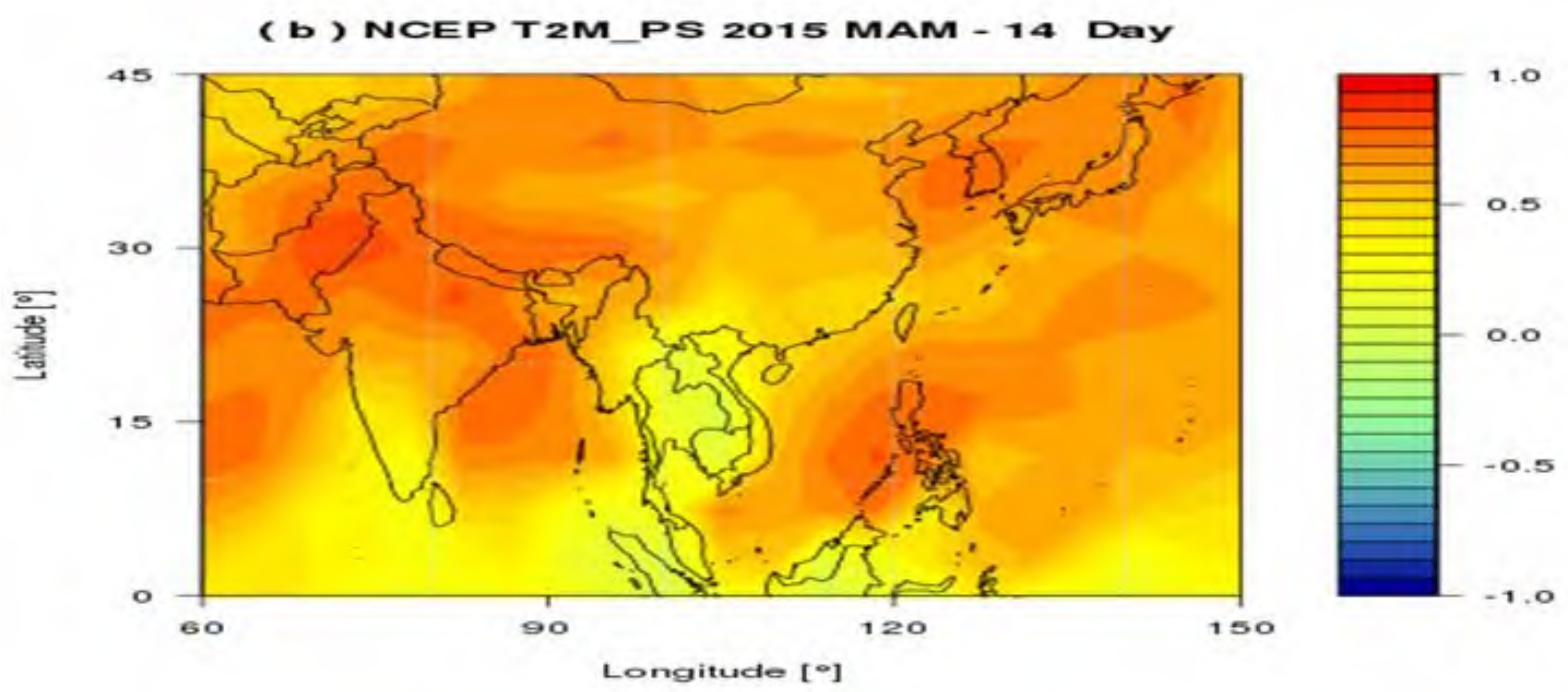
NCEP S2S Data

- Variables: **T2M**, **MSLP**, **TOTPRCP**
- Lead Times: 7 Days, 14 Days, 21 Days, 28 Days
- Study Area: (latitude: 0°N-45°N ; longitude:60°E-150°E)
- Study Seasons: 2015 Spring, 2015 Summer, 2015 Fall, 2015-2016 Winter, 2016 Spring

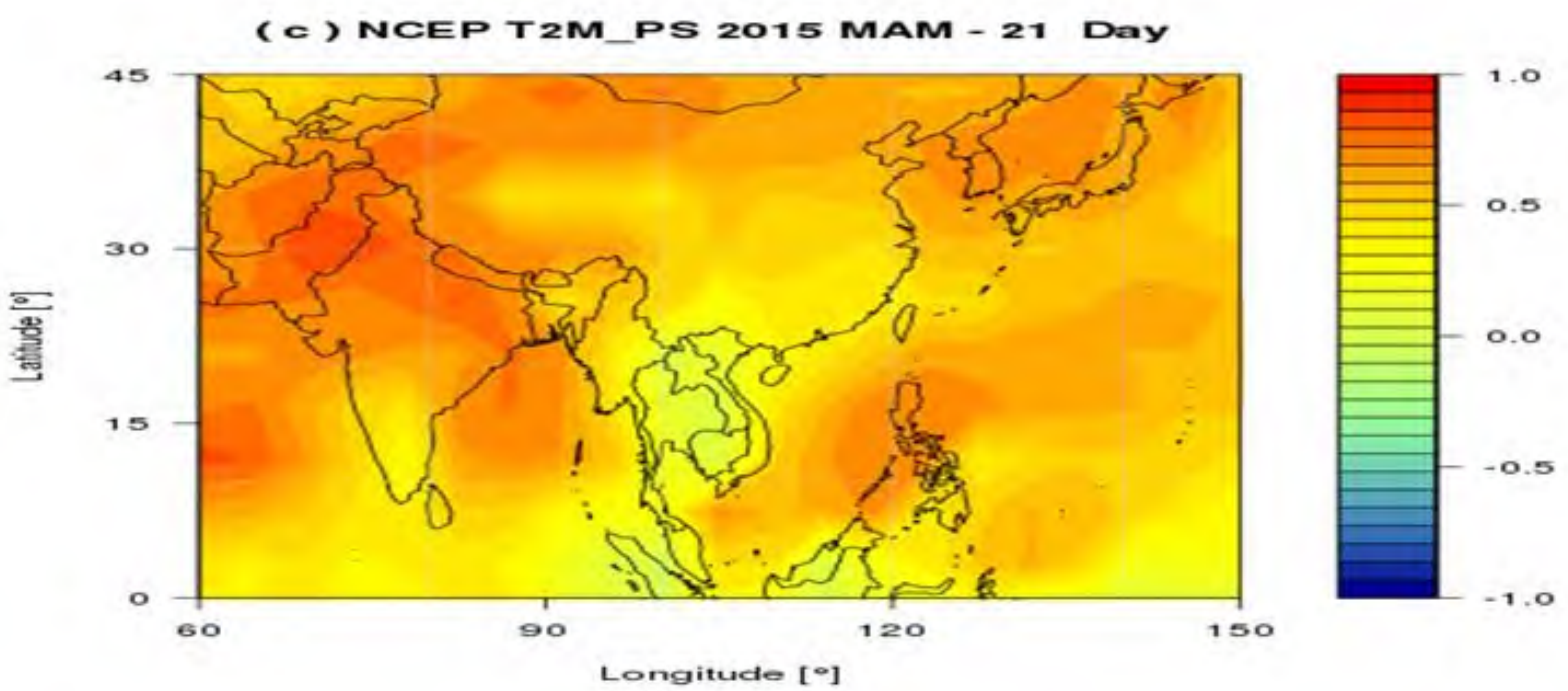
NCEP S2S 2015 Spring T2M PS



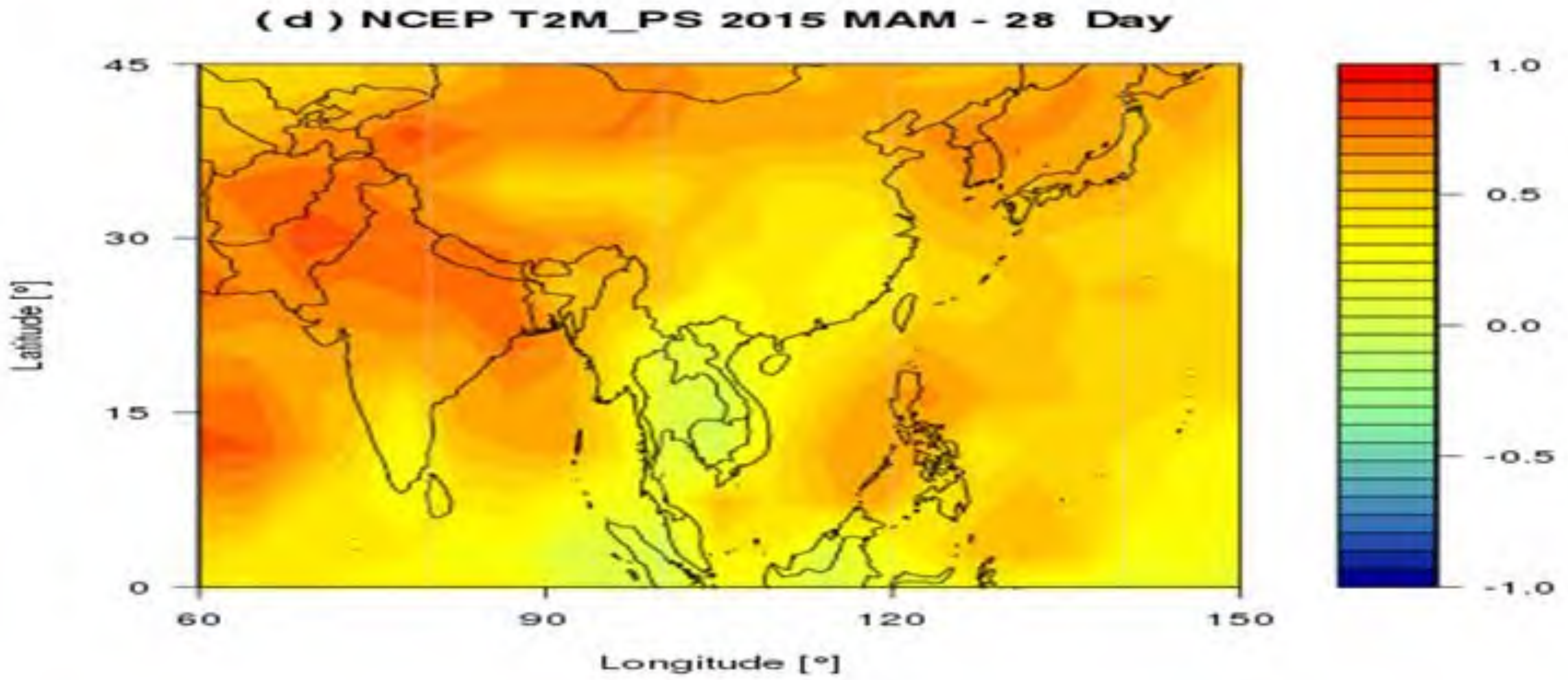
NCEP S2S 2015 Spring T2M PS



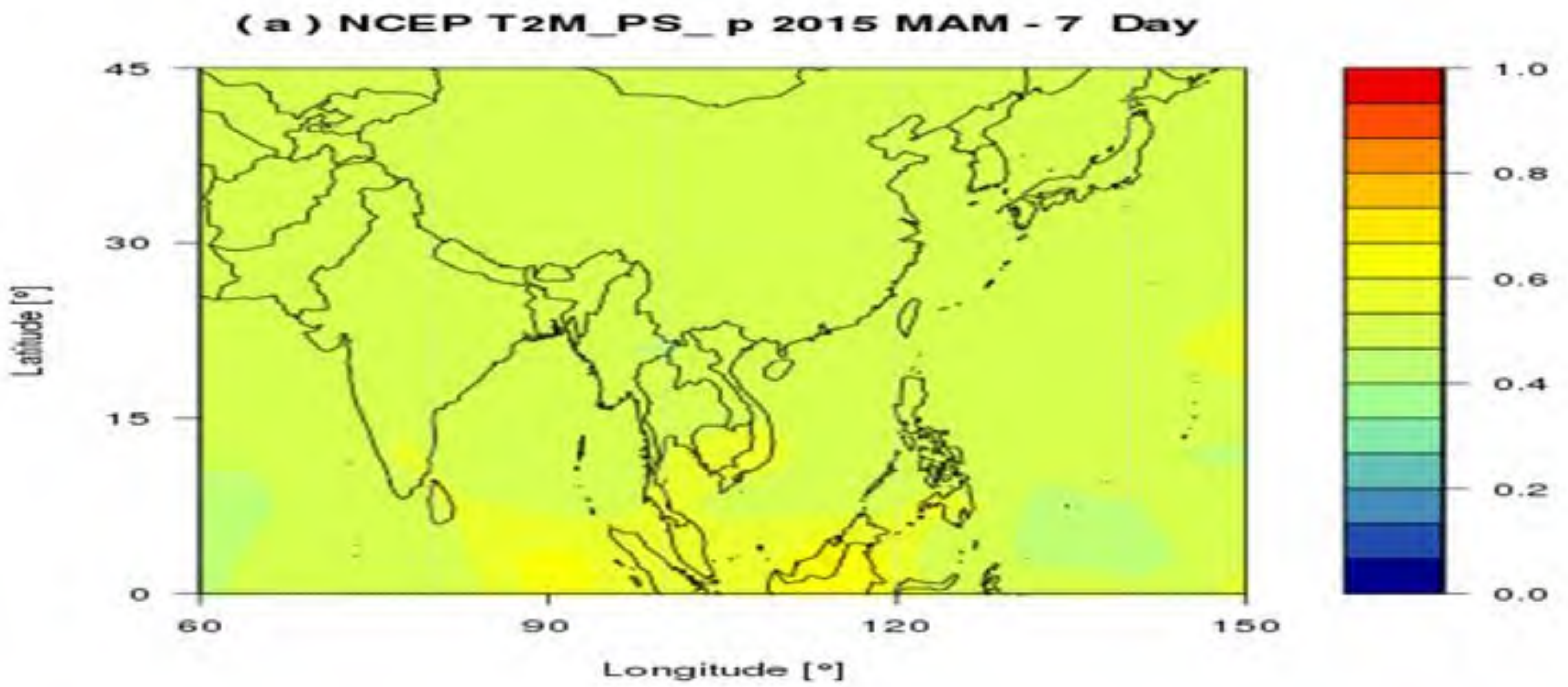
NCEP S2S 2015 Spring T2M PS



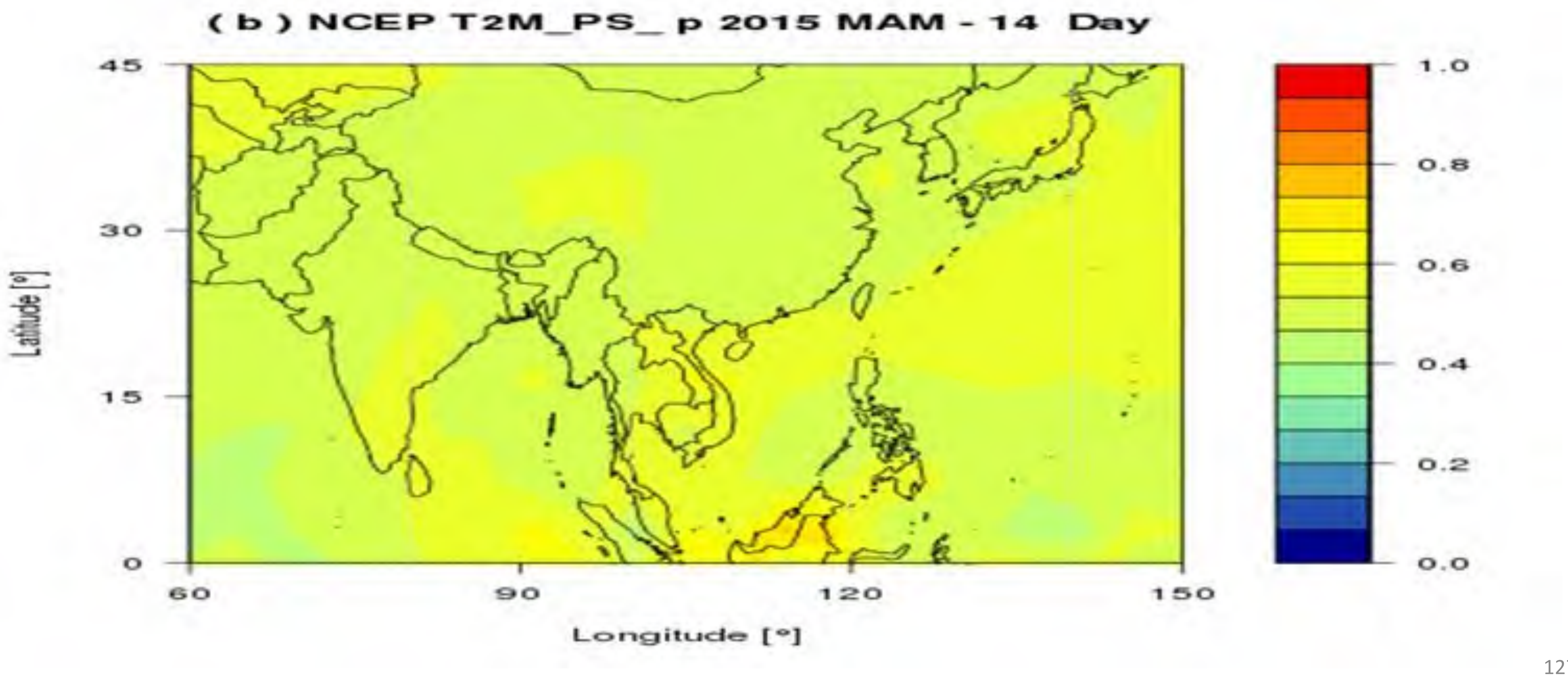
NCEP S2S 2015 Spring T2M PS



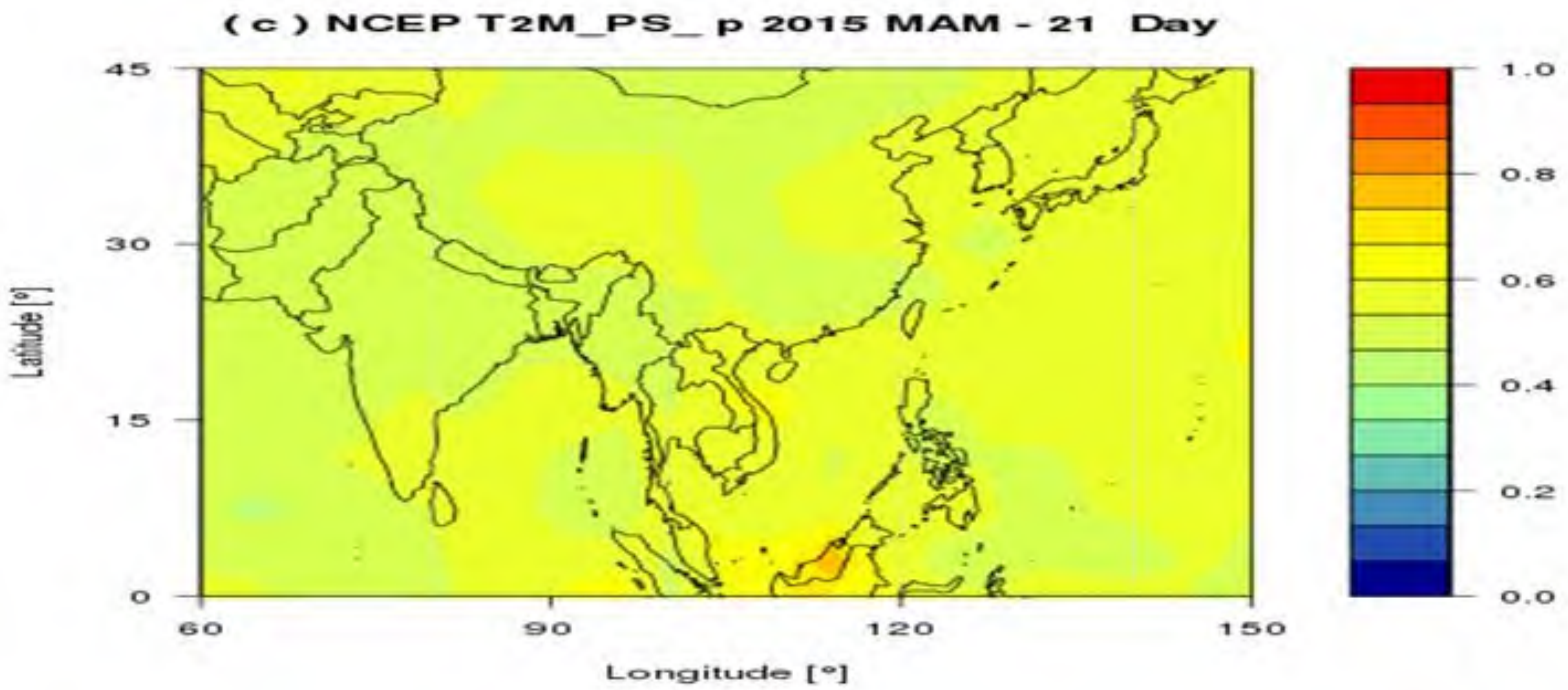
NCEP S2S 2015 Spring T2M PS_ $\bar{p}_Q$



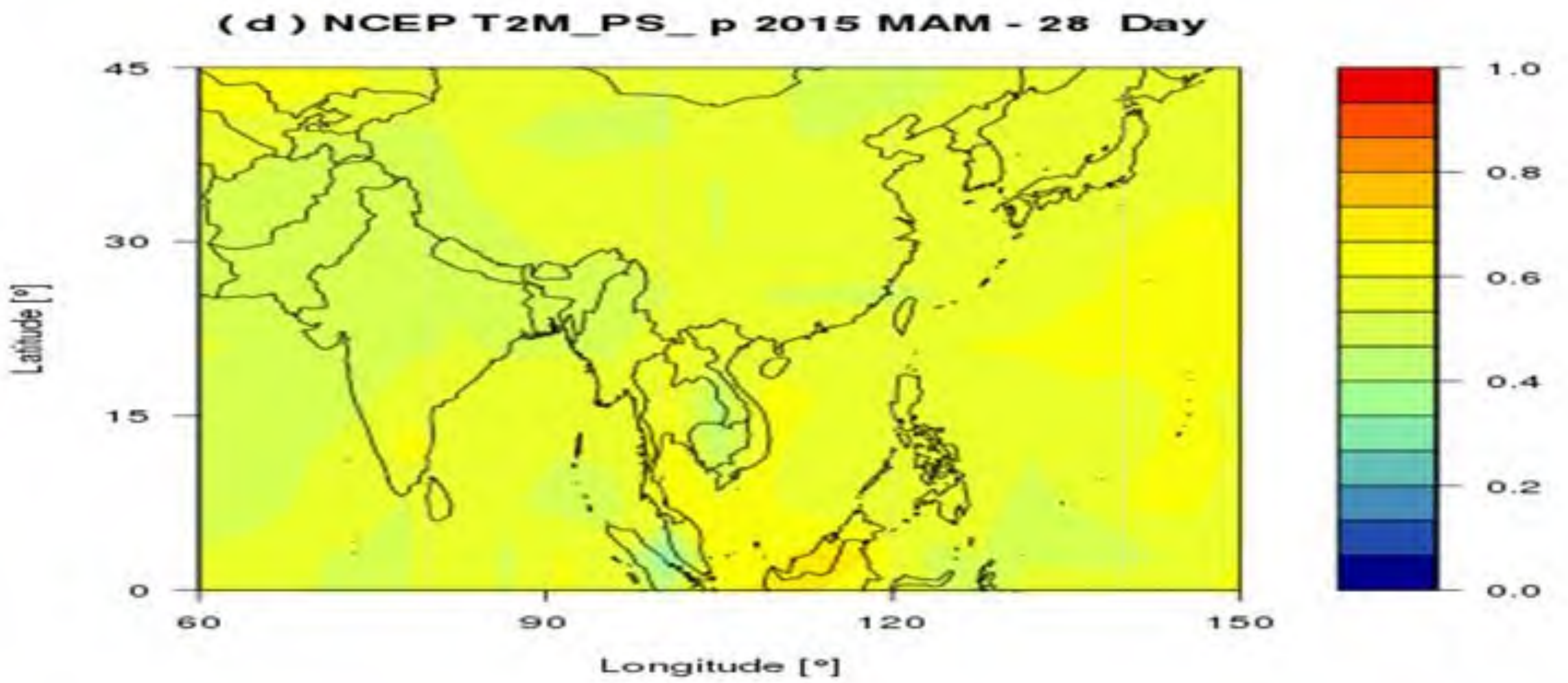
NCEP S2S 2015 Spring T2M PS_ $\bar{p}_Q$



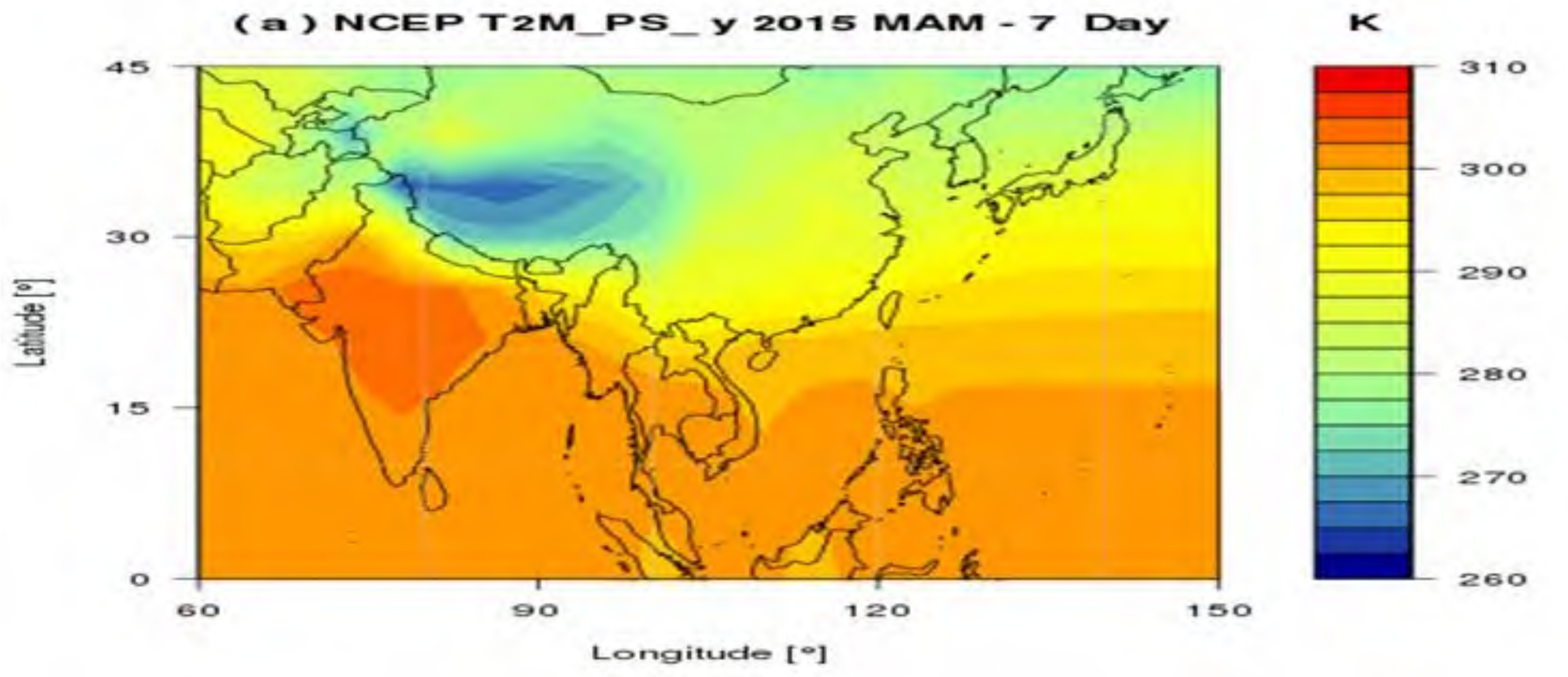
NCEP S2S 2015 Spring T2M PS_ $\bar{p}_Q$



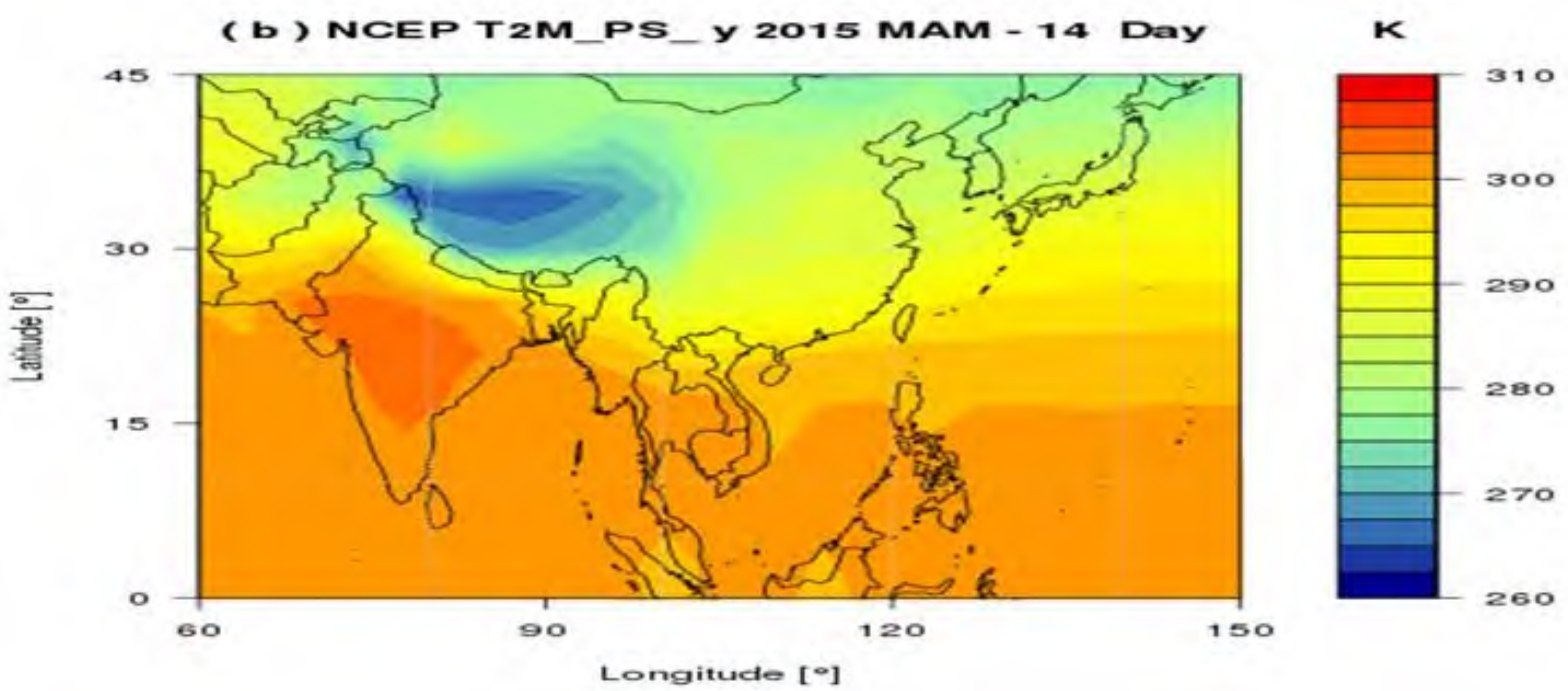
NCEP S2S 2015 Spring T2M PS_ $\bar{p}_Q$



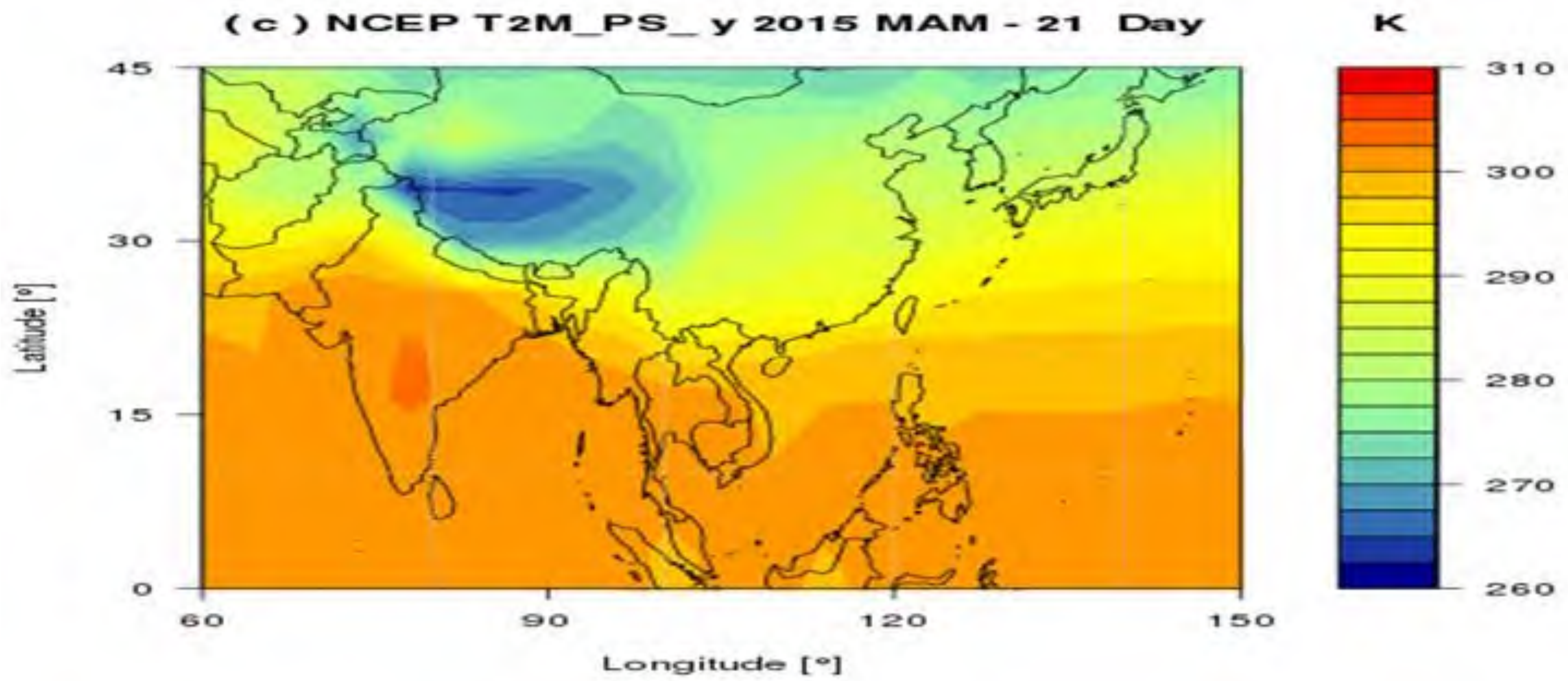
NCEP S2S 2015 Spring T2M PS_ $\bar{p}_Q$(y)



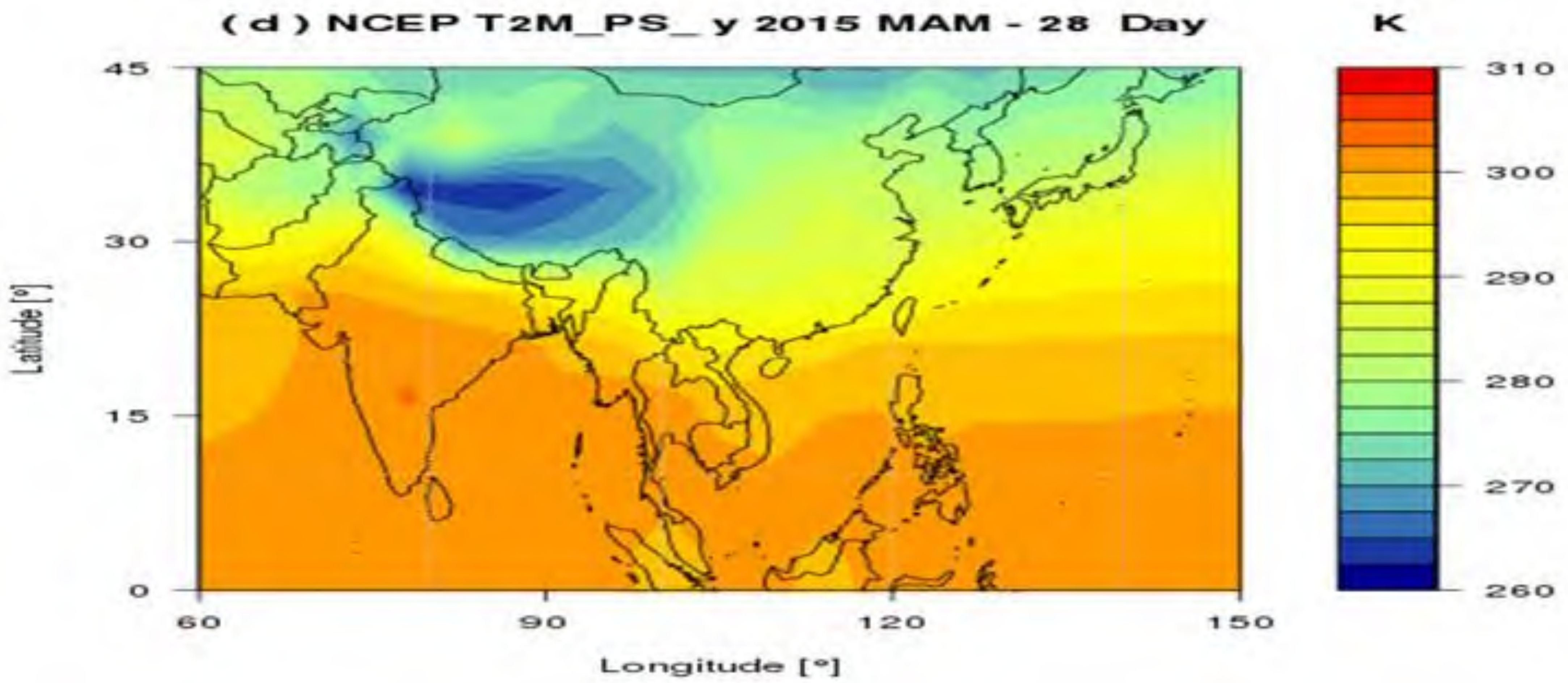
NCEP S2S 2015 Spring T2M PS_ $\bar{p}_Q$(y)



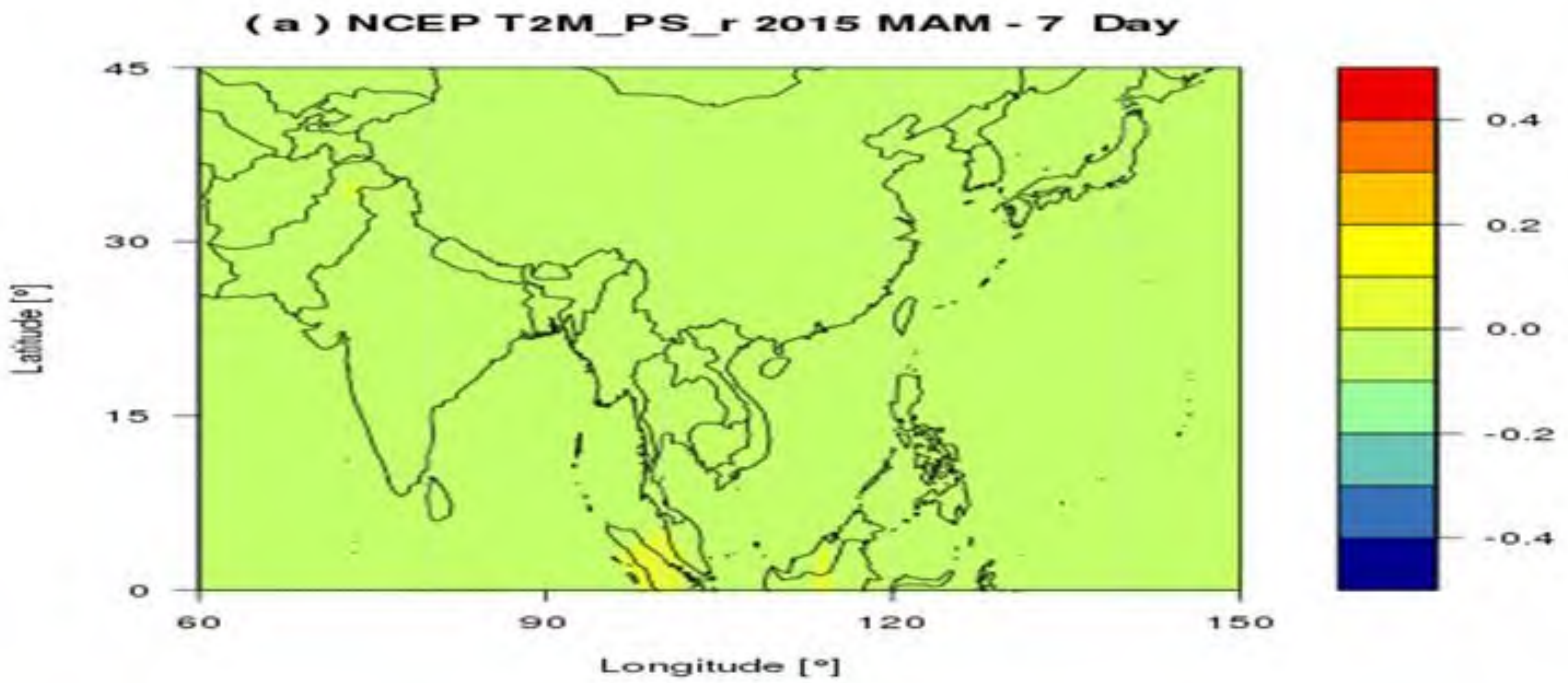
NCEP S2S 2015 Spring T2M PS_ $\bar{p}_Q$(y)



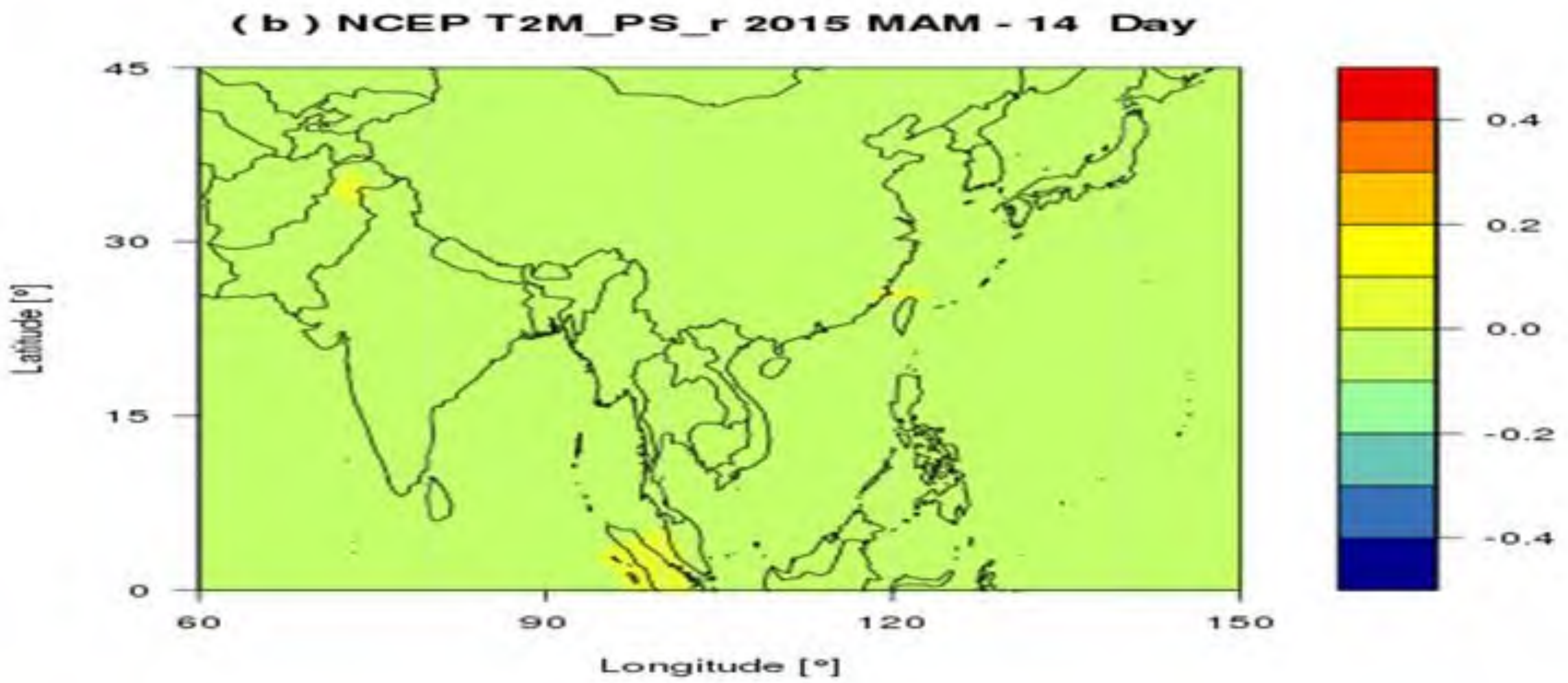
NCEP S2S 2015 Spring T2M PS_ $\bar{p}_Q$(y)



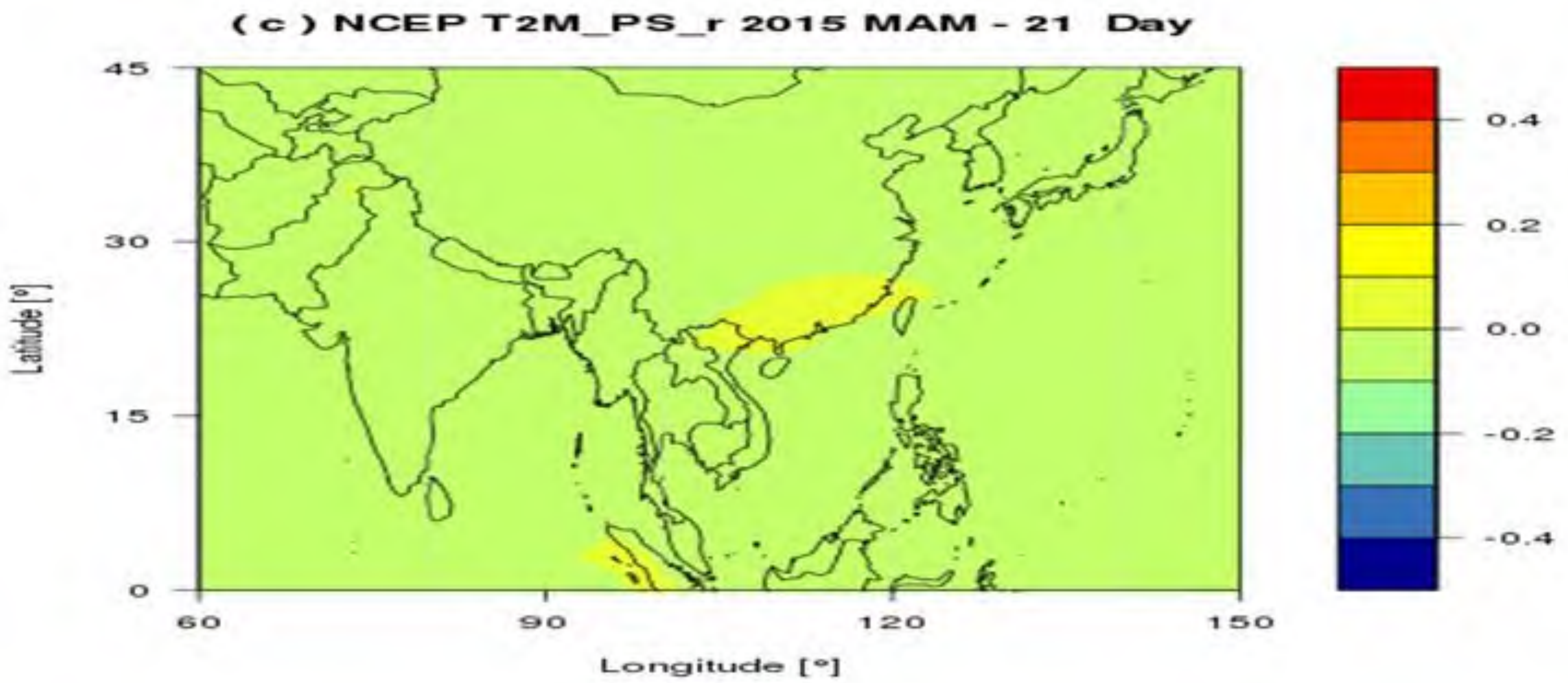
NCEP S2S 2015 Spring T2M PS_ γ_{Q_p}



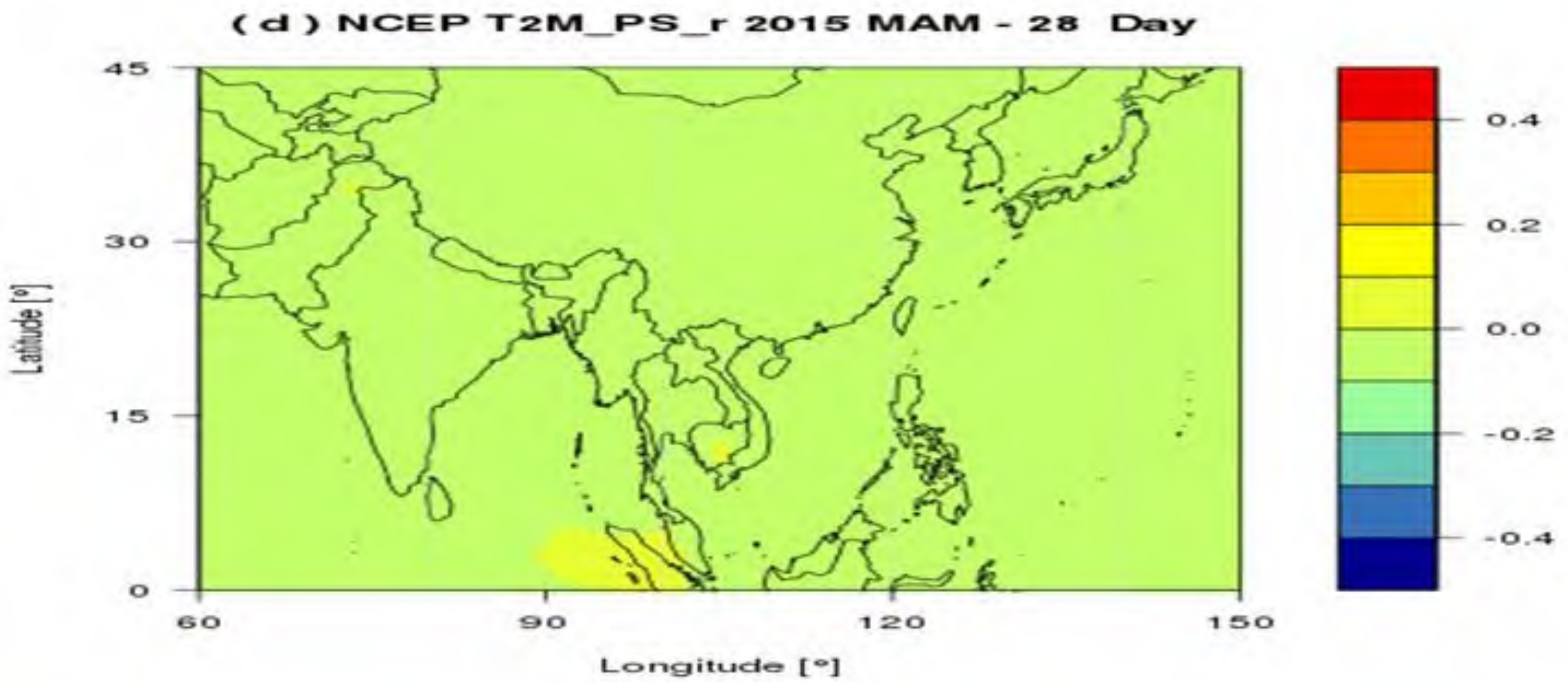
NCEP S2S 2015 Spring T2M PS_ γ_{Q_p}



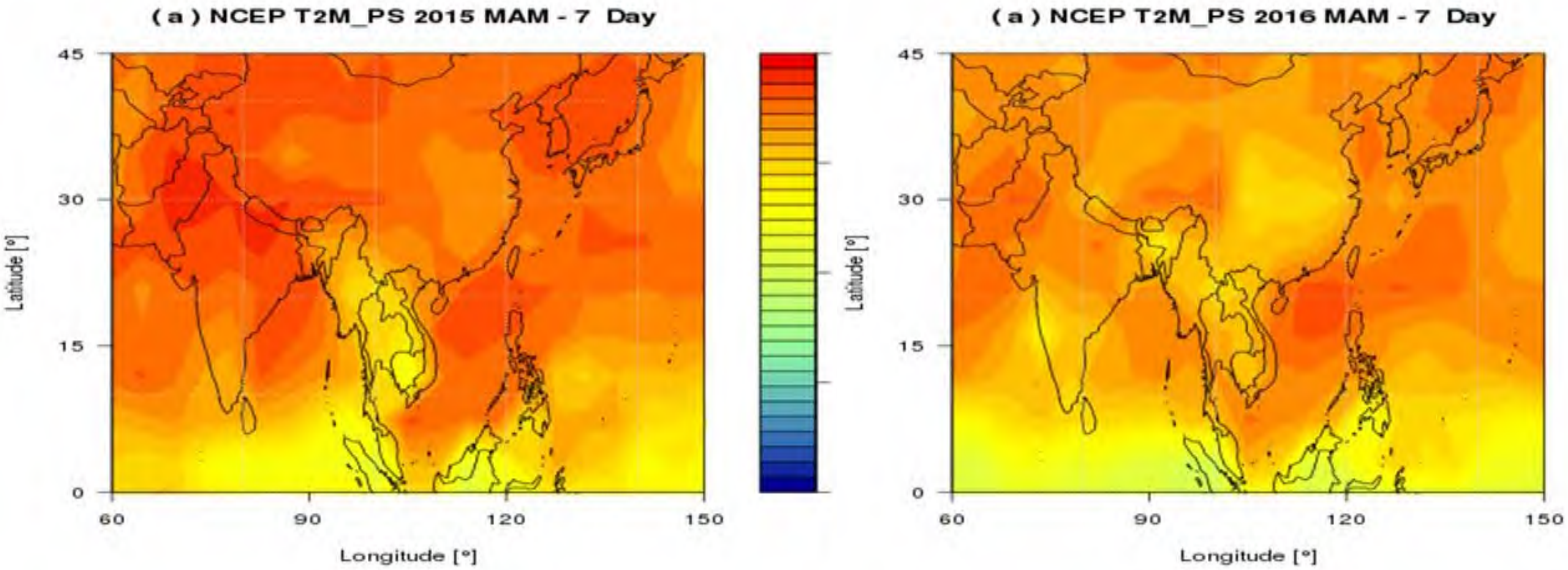
NCEP S2S 2015 Spring T2M PS_ γ_{Q_p}



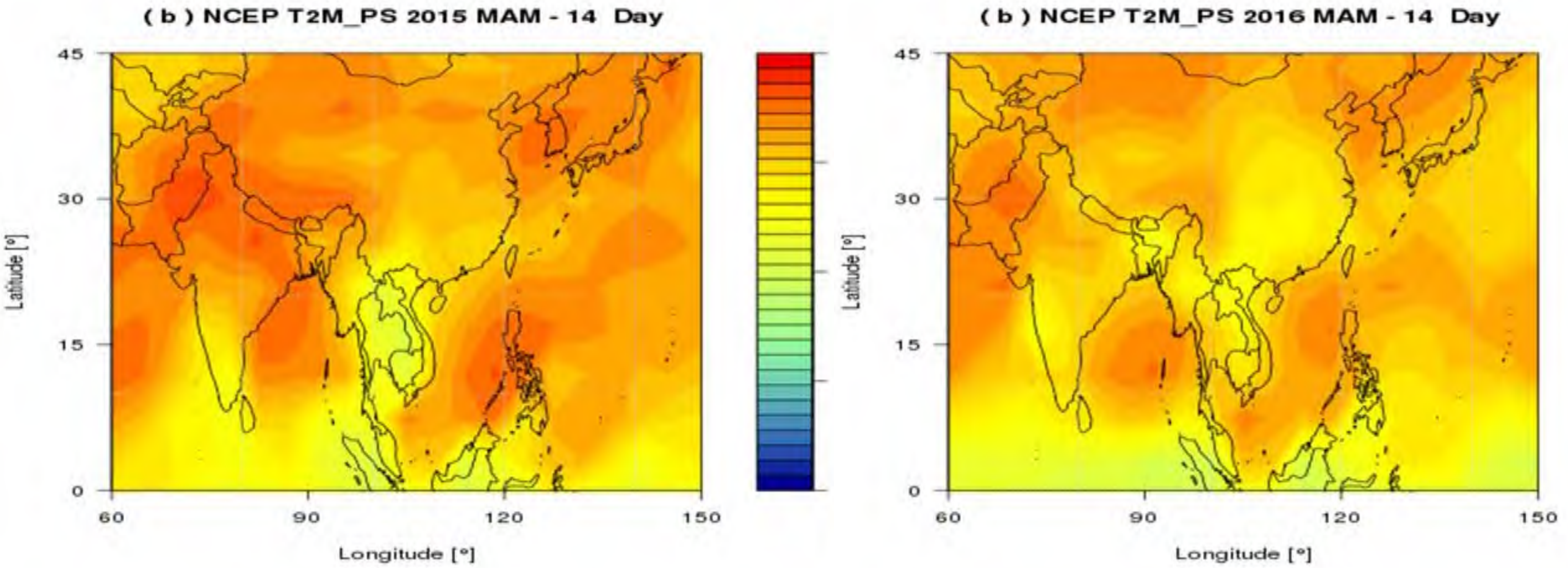
NCEP S2S 2015 Spring T2M PS_ γ_{Q_p}



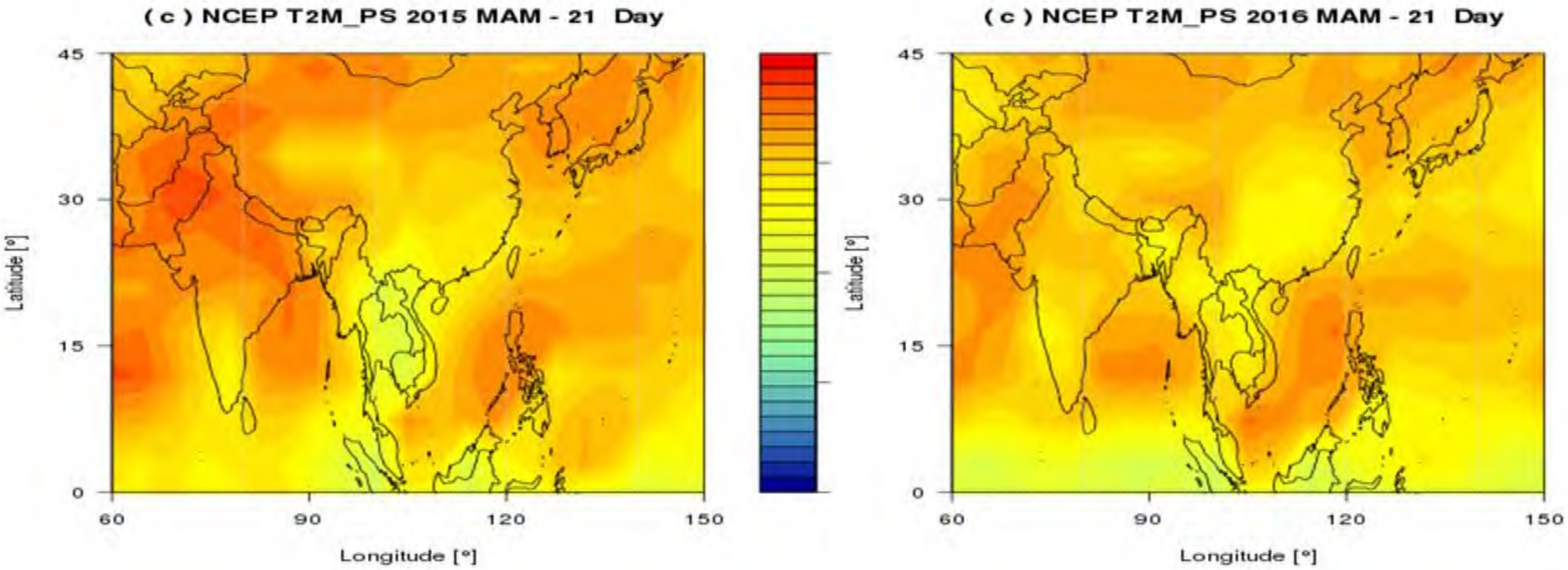
NCEP S2S 2015 VS 2016 Spring T2M PS



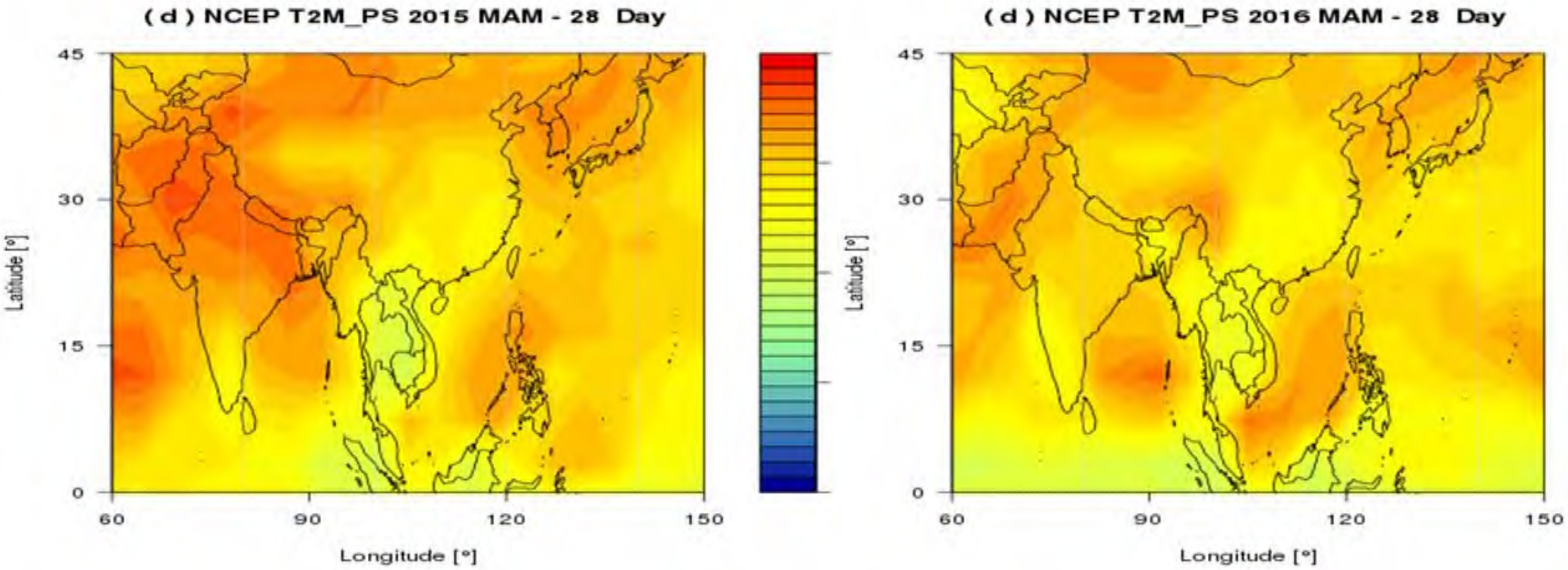
NCEP S2S 2015 VS 2016 Spring T2M PS



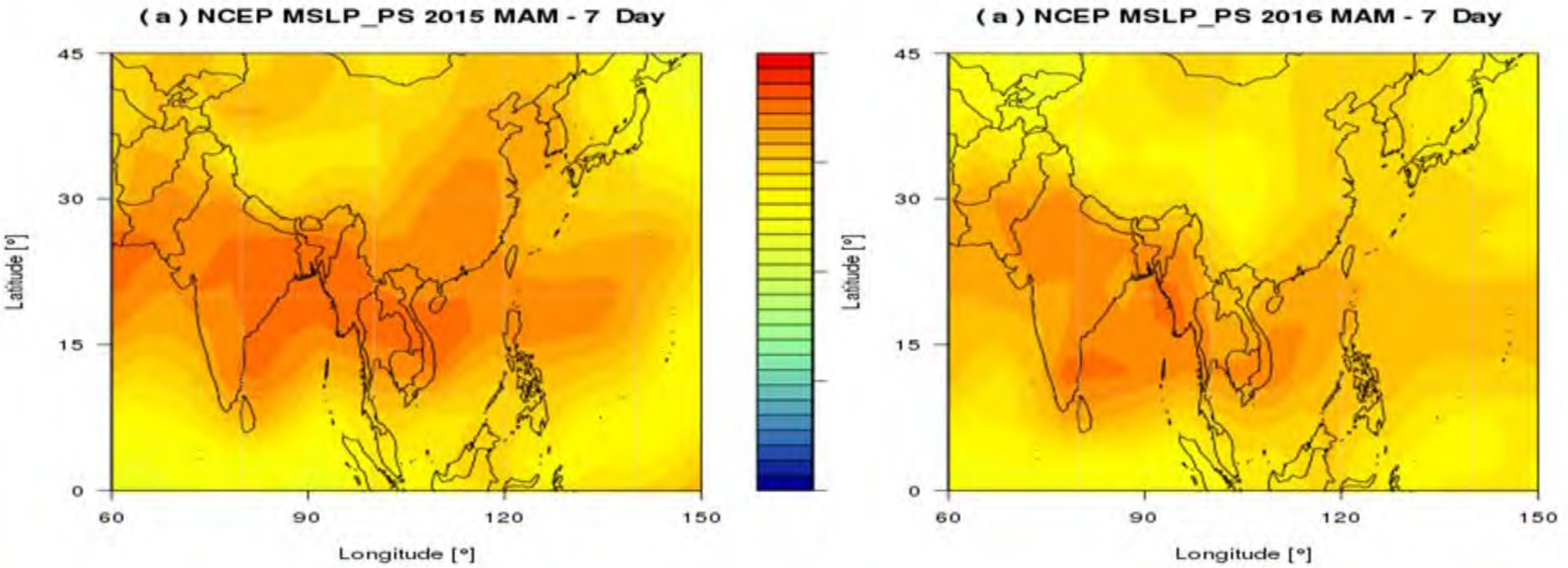
NCEP S2S 2015 VS 2016 Spring T2M PS



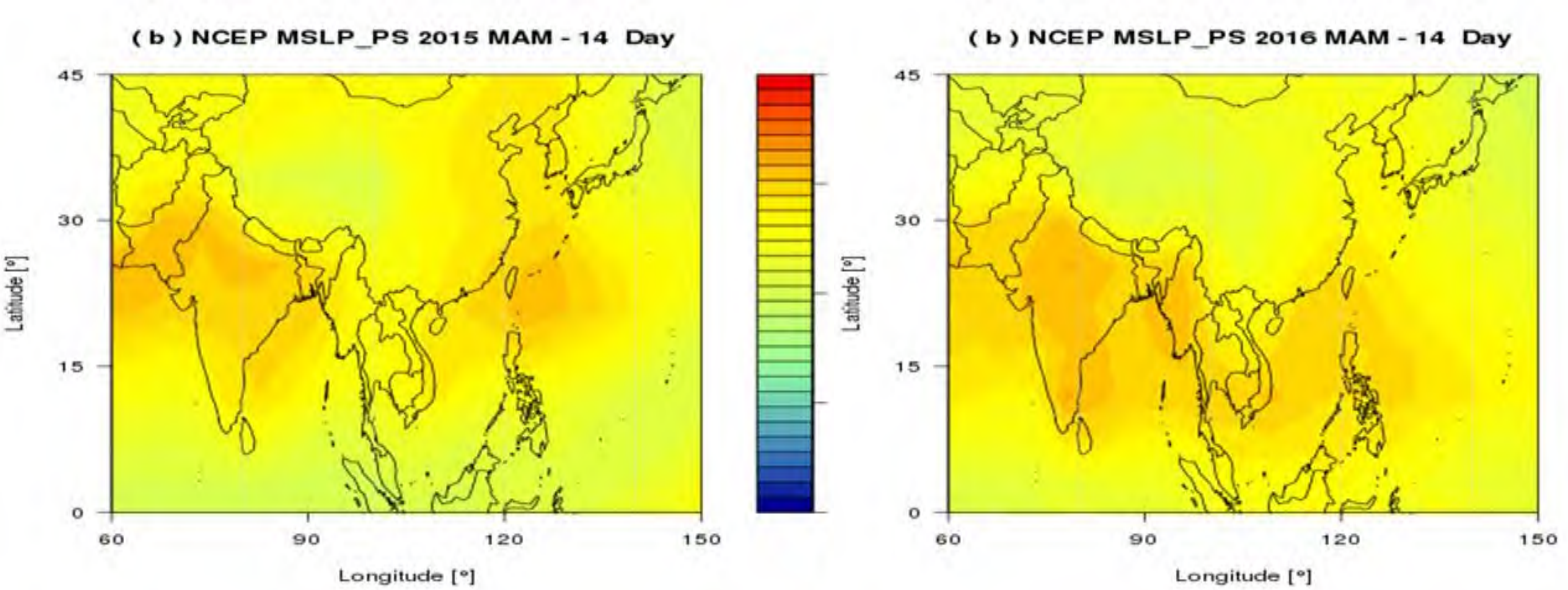
NCEP S2S 2015 VS 2016 Spring T2M PS



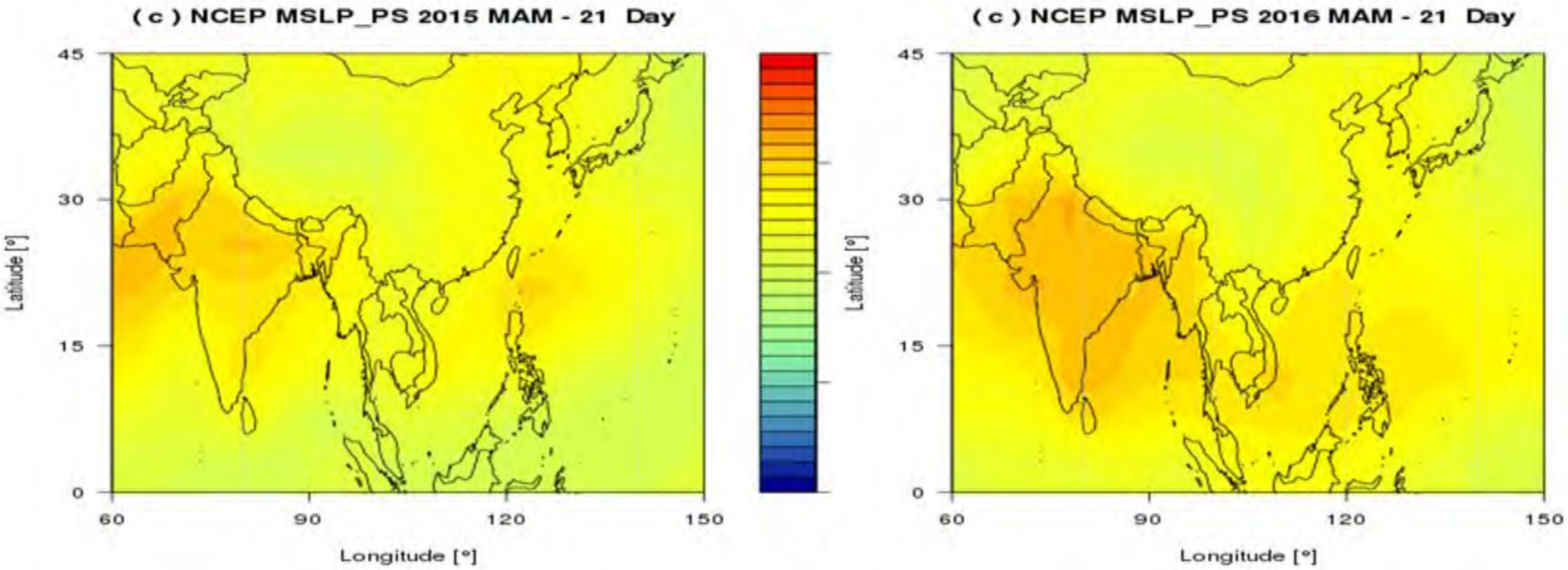
NCEP S2S 2015 VS 2016 Spring MSLP PS



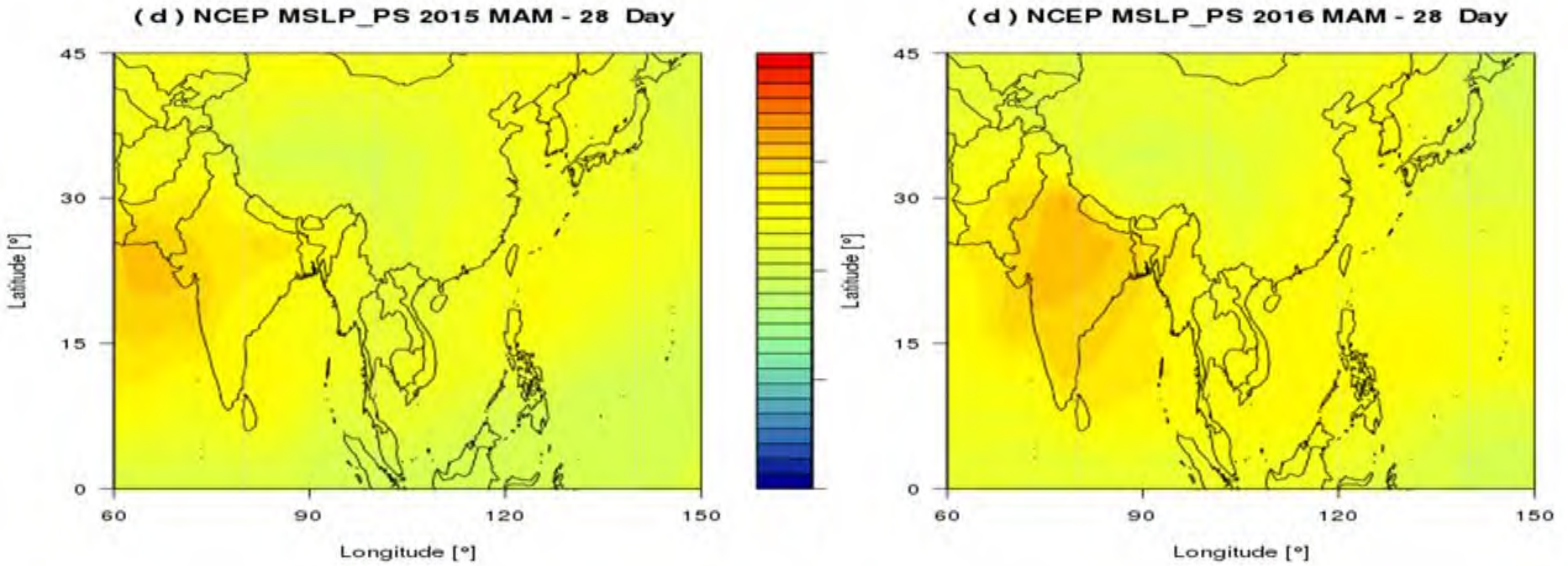
NCEP S2S 2015 VS 2016 Spring MSLP PS



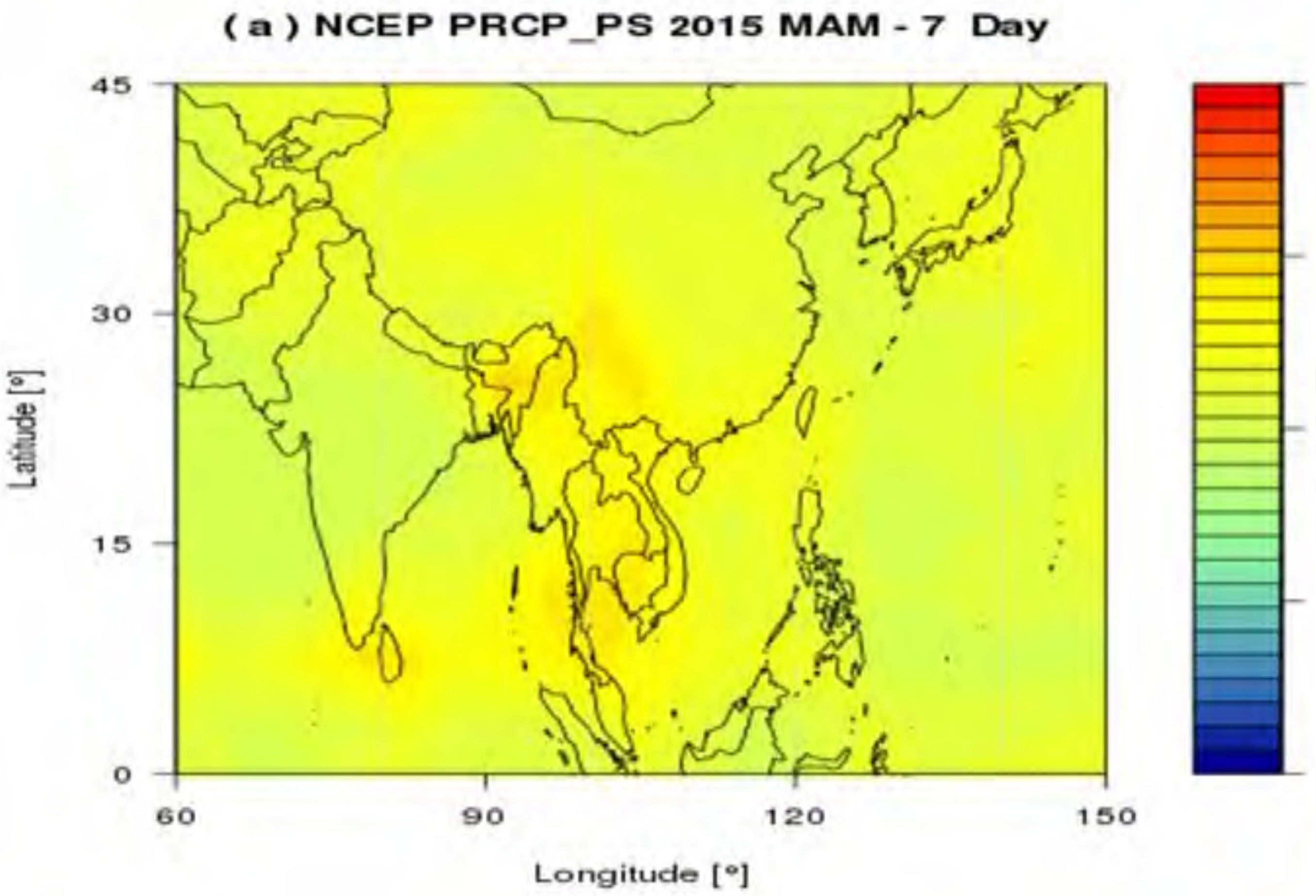
NCEP S2S 2015 VS 2016 Spring MSLP PS

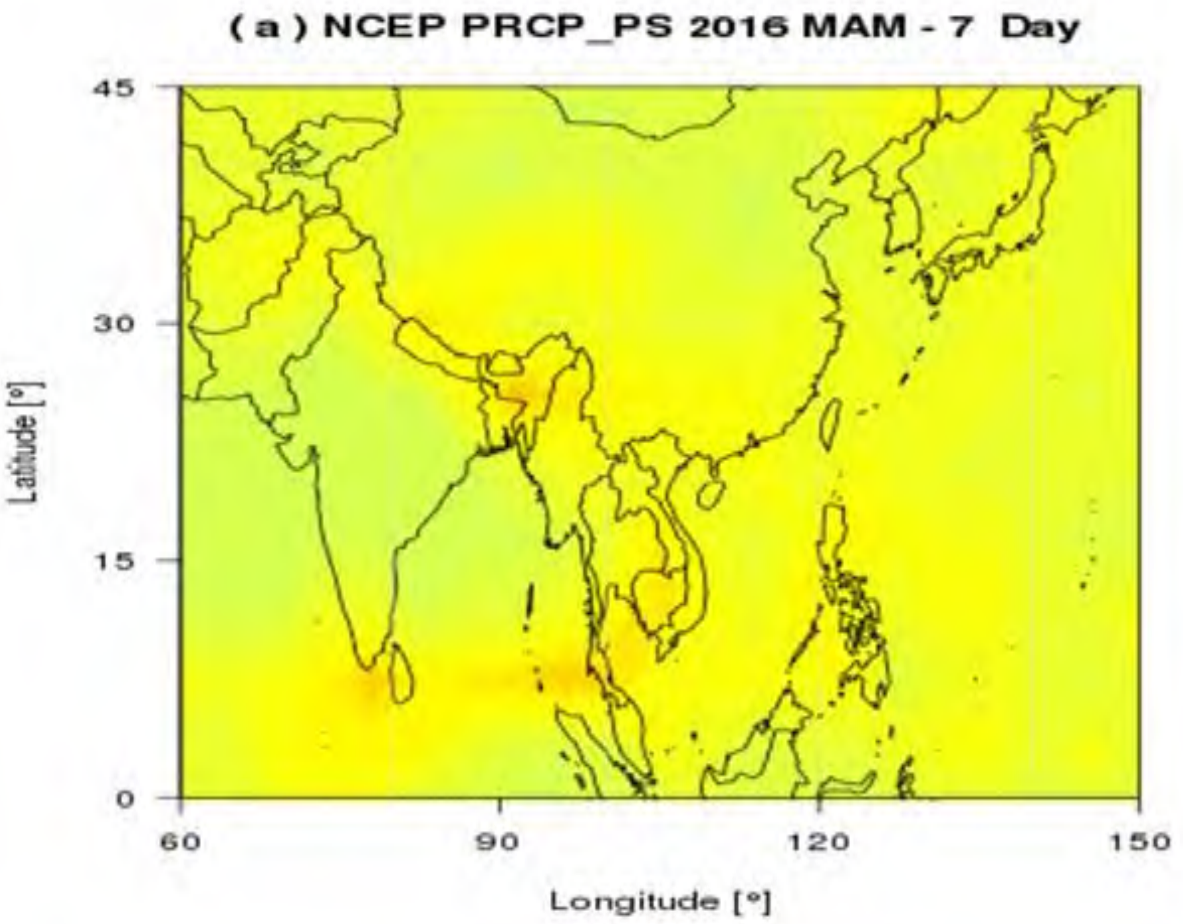


NCEP S2S 2015 VS 2016 Spring MSLP PS

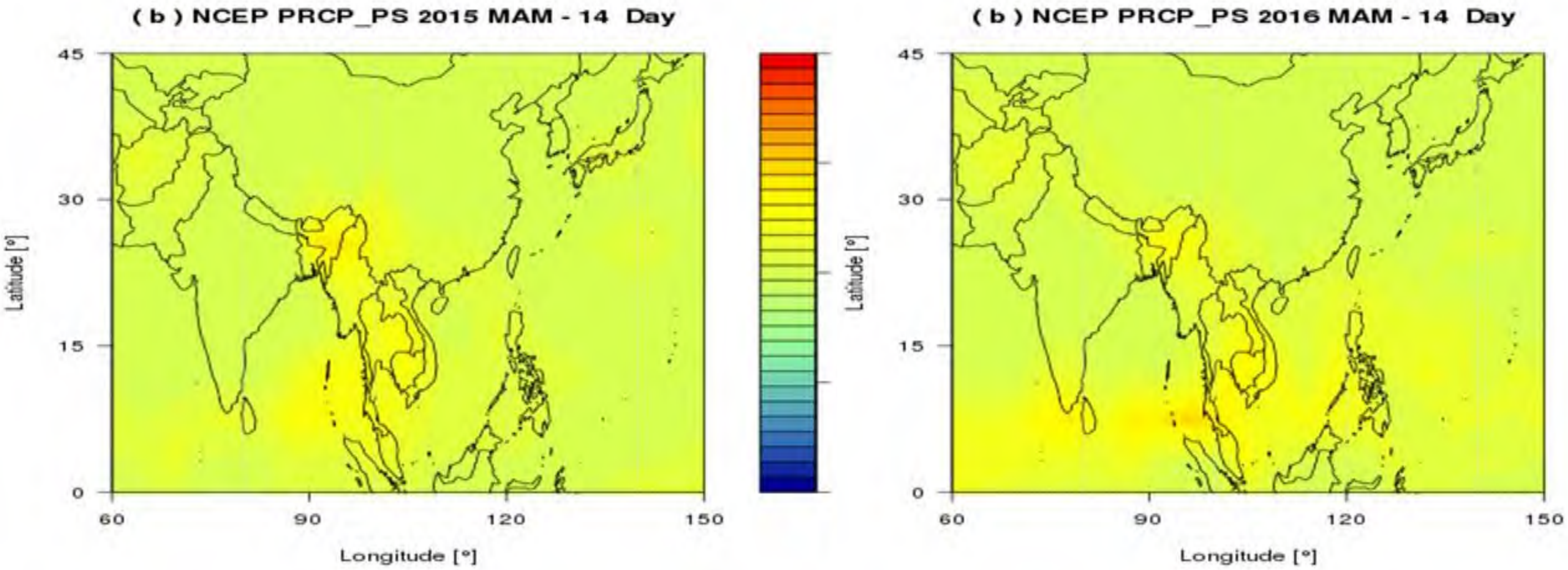


NCEP S2S 2015 VS 2016 Spring TOTPRCP PS

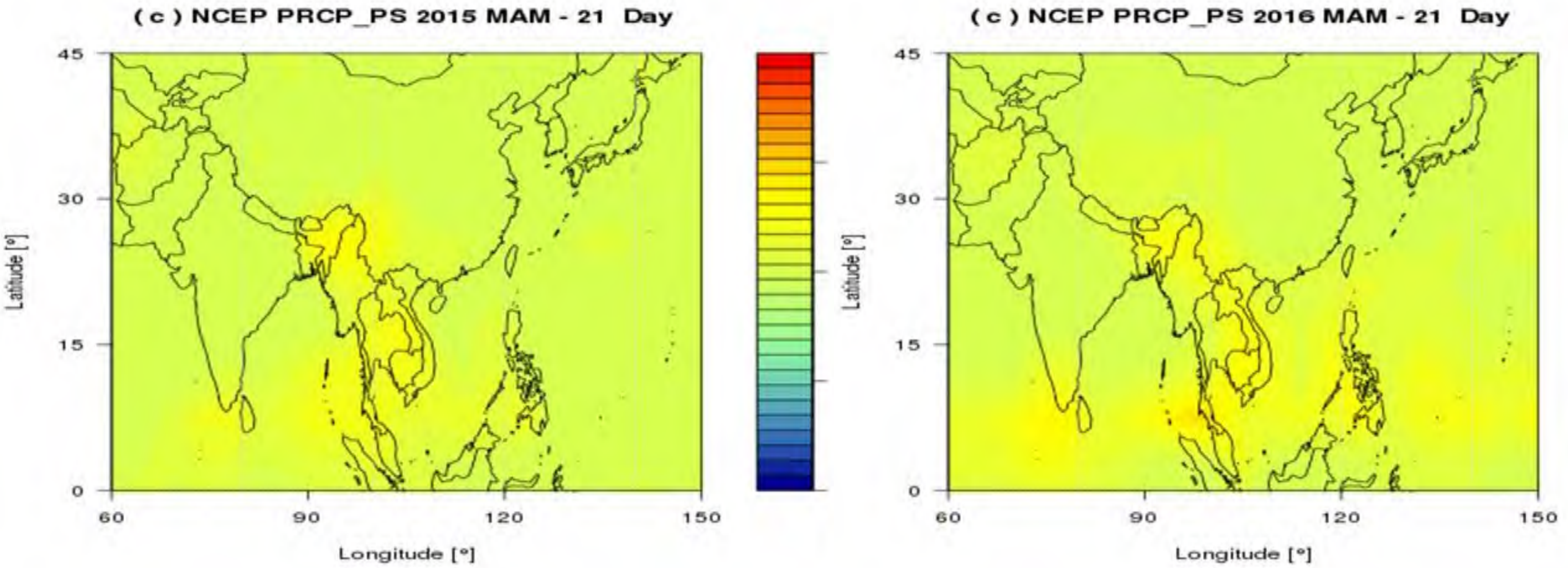




NCEP S2S 2015 VS 2016 Spring TOTPRCP PS



NCEP S2S 2015 VS 2016 Spring TOTPRCP PS



NCEP S2S 2015 VS 2016 Spring TOTPRCP PS