# Can SMYLE Beat My First Best Guess of the PDO?

## Accounting for Inter-model Correlation when Testing for Model Skill Differences

**Evan Meeker[1]**

Elizabeth Maroon[1], Anna-Lena Deppenmeier[2], LuAnne Thompson[3], Daniel Vimont[1], Stephen Yeager[4]

[1]Department of Atmospheric and Oceanic Sciences, University of Wisconsin Madison

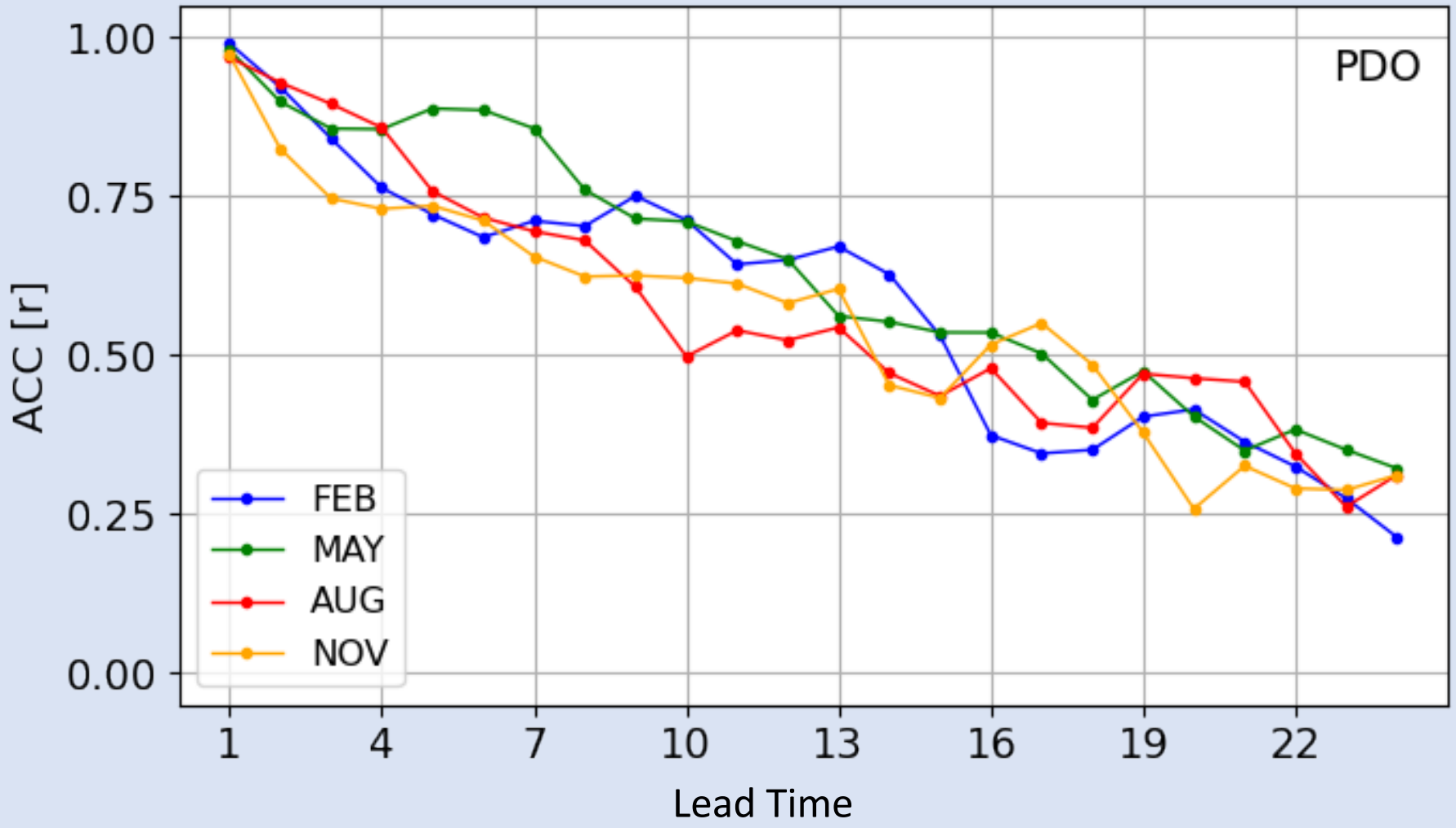
[2]Department of Earth, Ocean and Ecological Sciences, University of Liverpool

[3]School of Oceanography, University of Washington Seattle

[4]National Center for Atmospheric Research, Boulder, Colorado

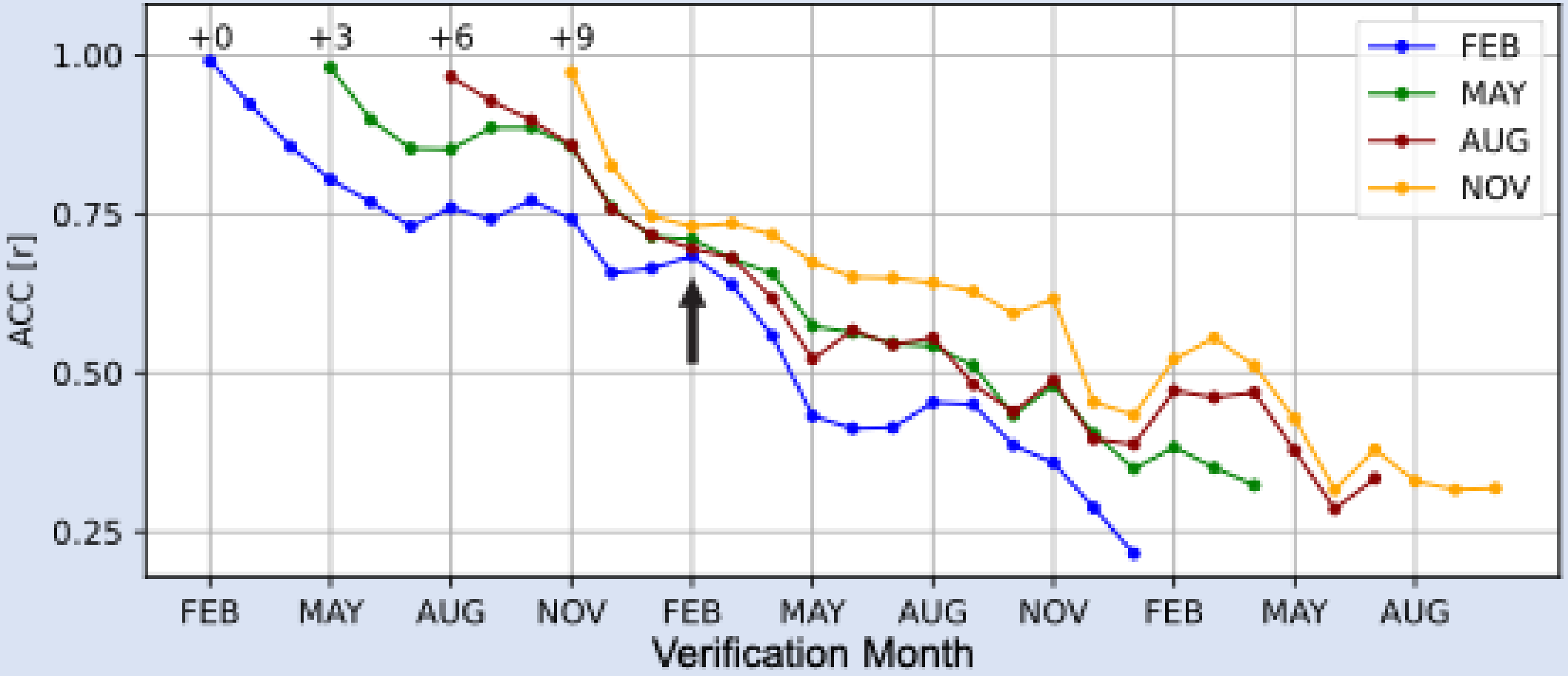
Work supported by the National Science Foundation under Grant OCE-2022740 and the National Defense Science and Engineering Graduate Fellowship

# SMYLE Pacific Decadal Oscillation (PDO) predictions have seasonality




*Meeker et al. (in revision at GRL)*

# SMYLE Pacific Decadal Oscillation (PDO) predictions have seasonality




*Meeker et al. (in revision at GRL)*

# Can SMYLE beat an AR1 Model fit to the observed PDO?

First-order Auto-Regressive Model (AR1):

$$P_t = \alpha P_{t-1} + \eta_t$$

Predicted PDO state at time *t*

PDO Lag-1 autocorrelation
$\alpha_{PDO} = 0.91$

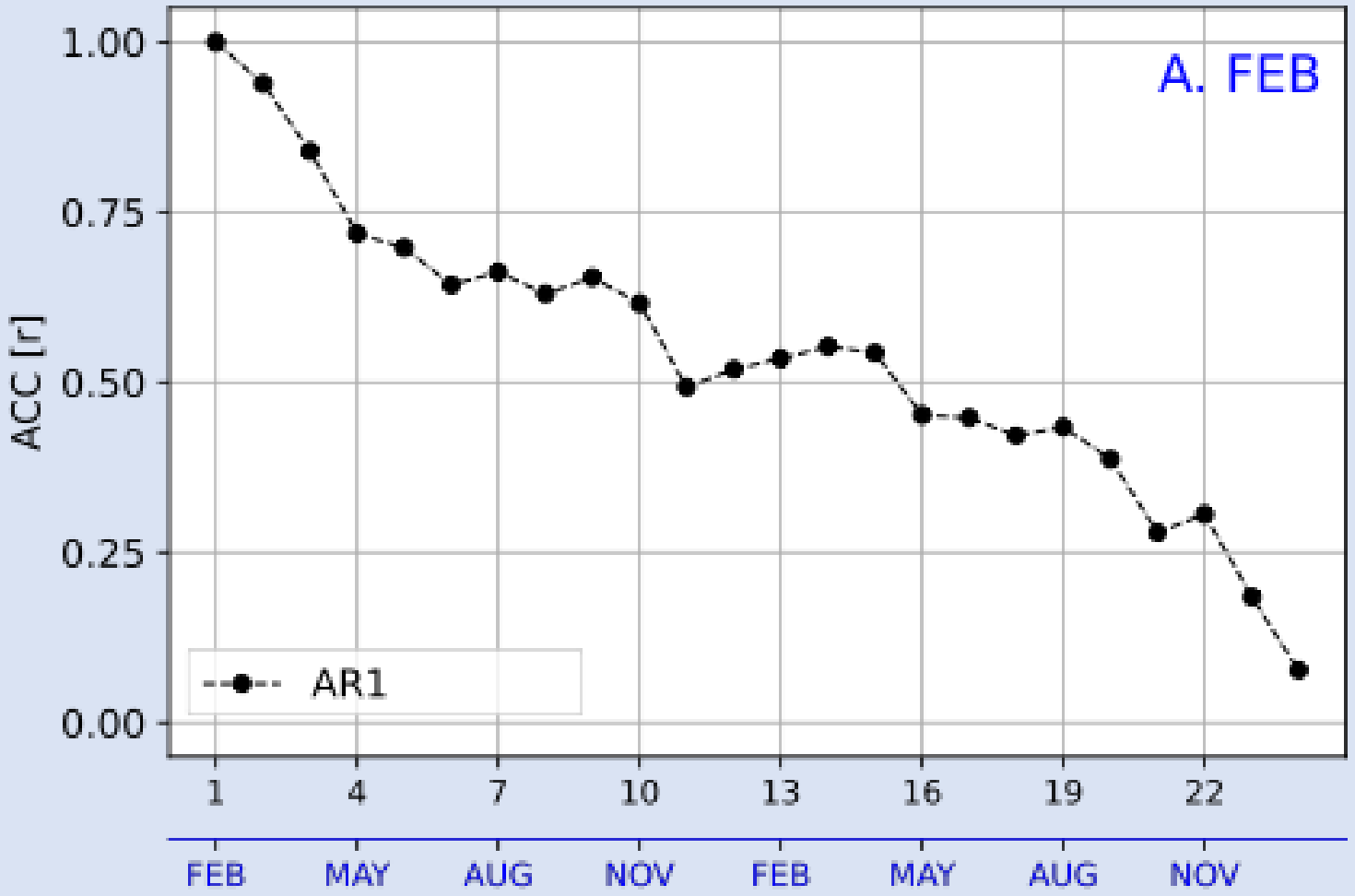
Previous PDO state at time *t-1*

White noise

# Can SMYLE beat an AR1 Model fit to the observed PDO?

First-order Auto-Regressive Model (AR1):

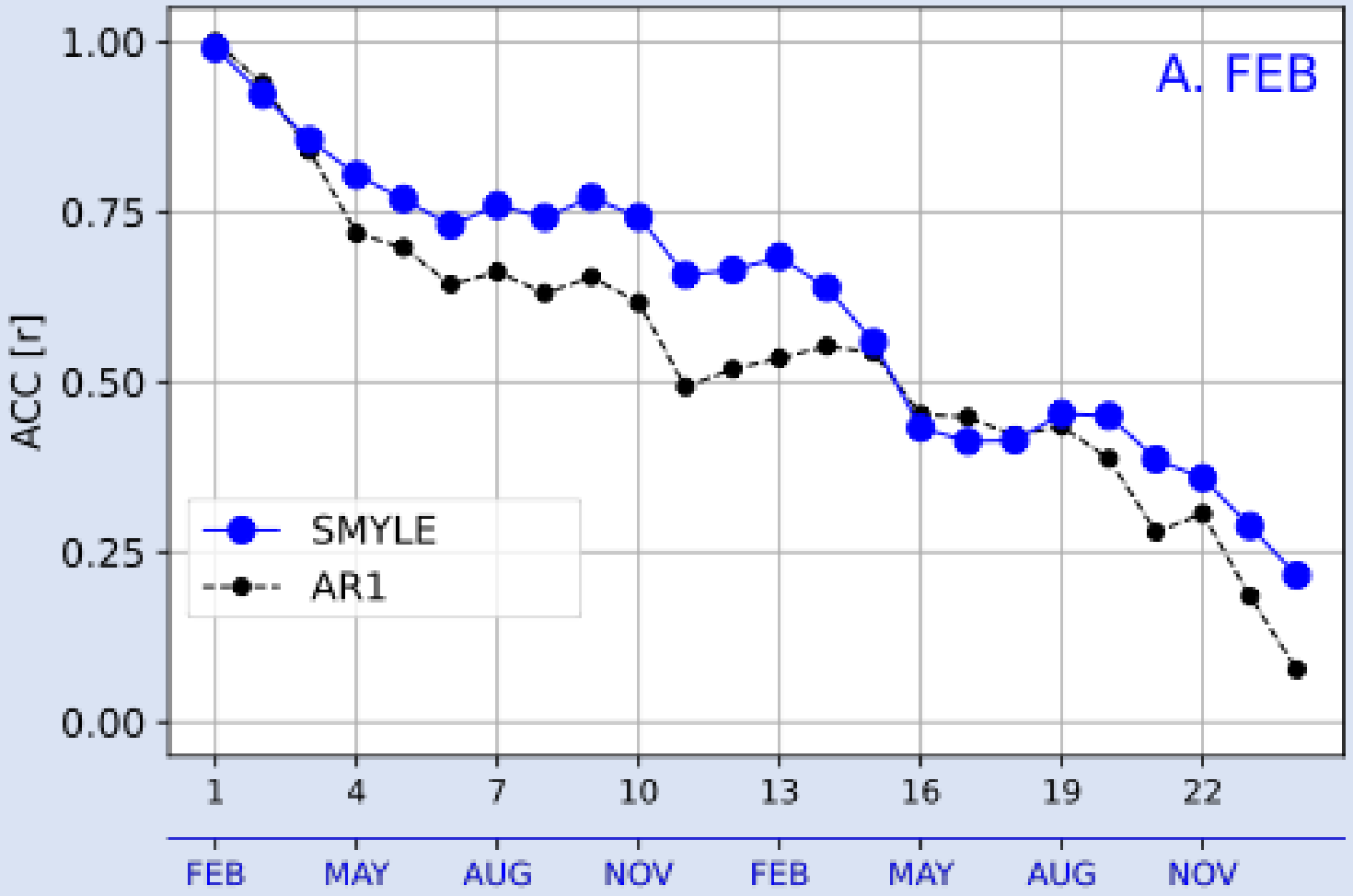
$$P_t = \alpha P_{t-1} + \eta_t$$



# Can SMYLE beat an AR1 Model fit to the observed PDO?

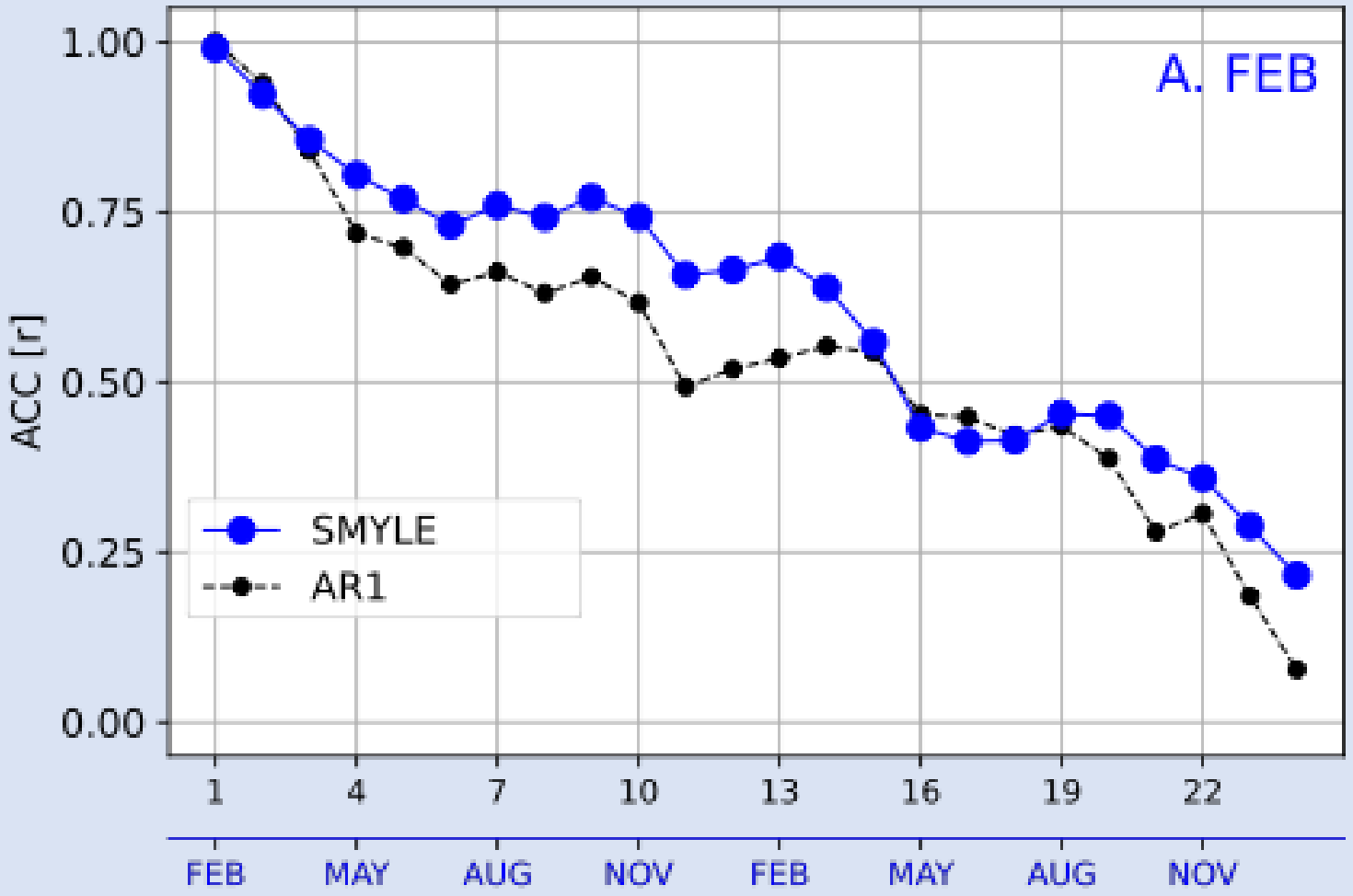
First-order Auto-Regressive Model (AR1):

$$P_t = \alpha P_{t-1} + \eta_t$$

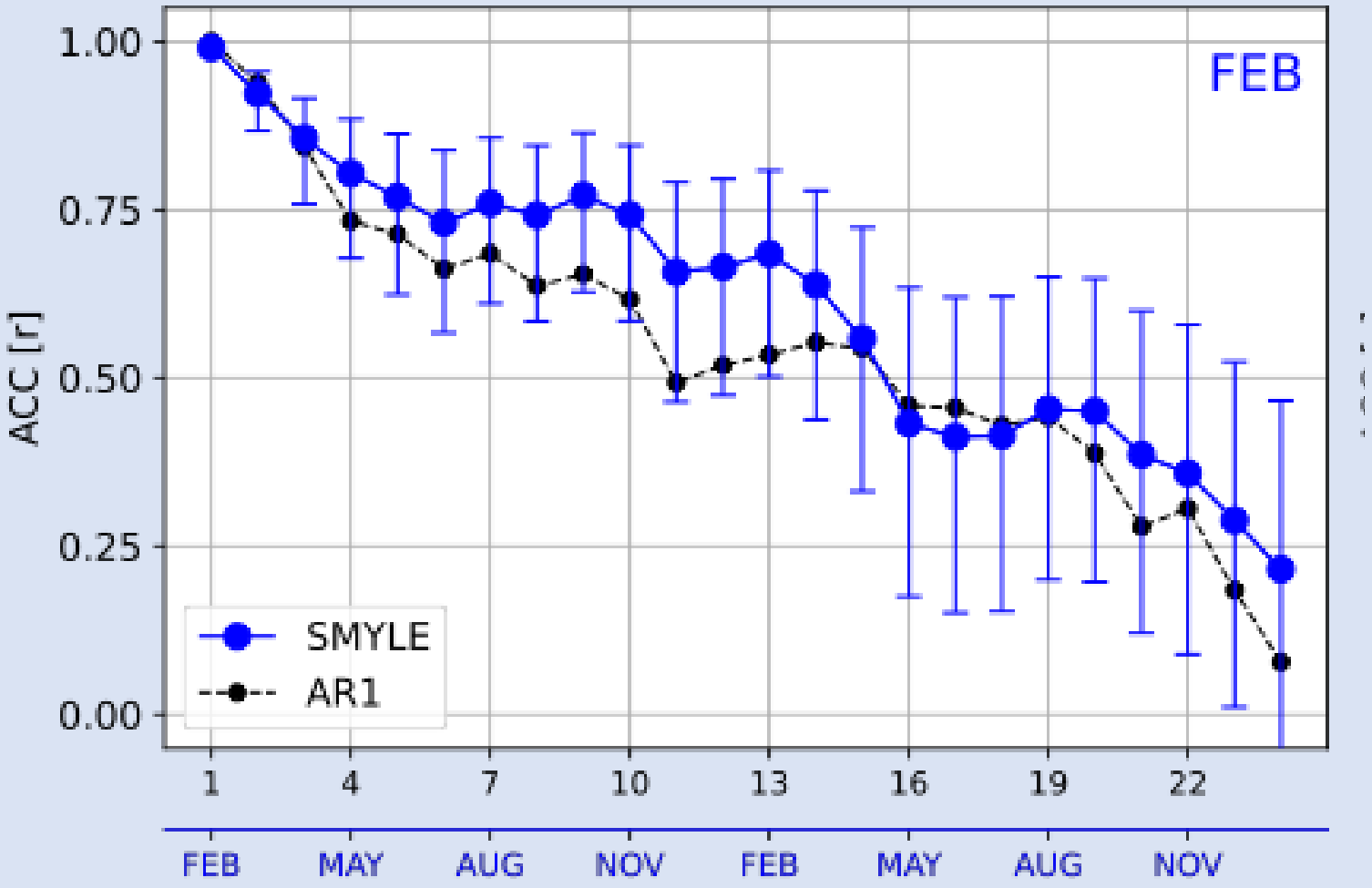


# Can SMYLE beat an AR1 Model fit to the observed PDO?

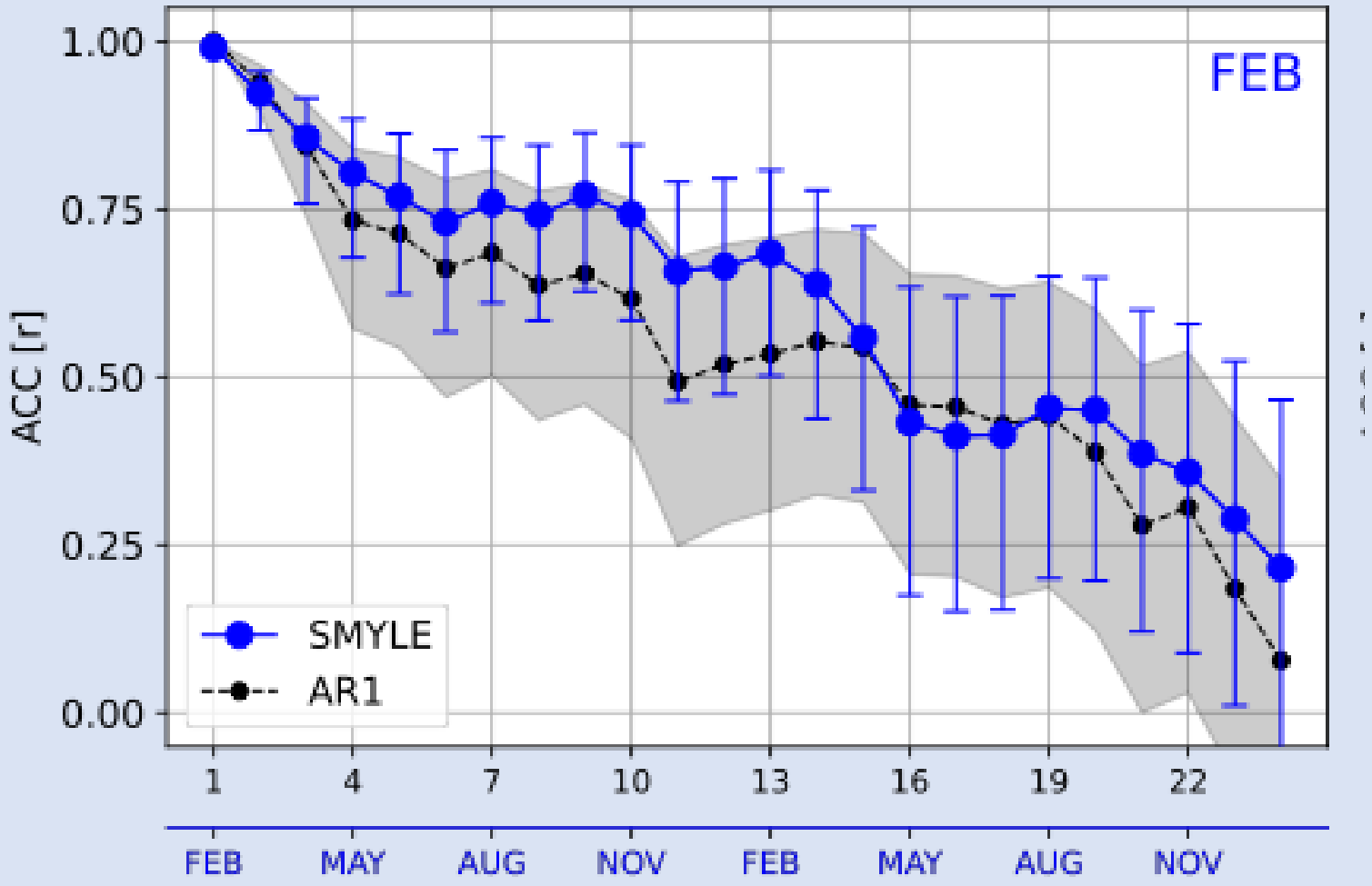
SMYLE skill is better. But is it ***significantly*** better?



# SMYLE uncertainty using standard confidence intervals

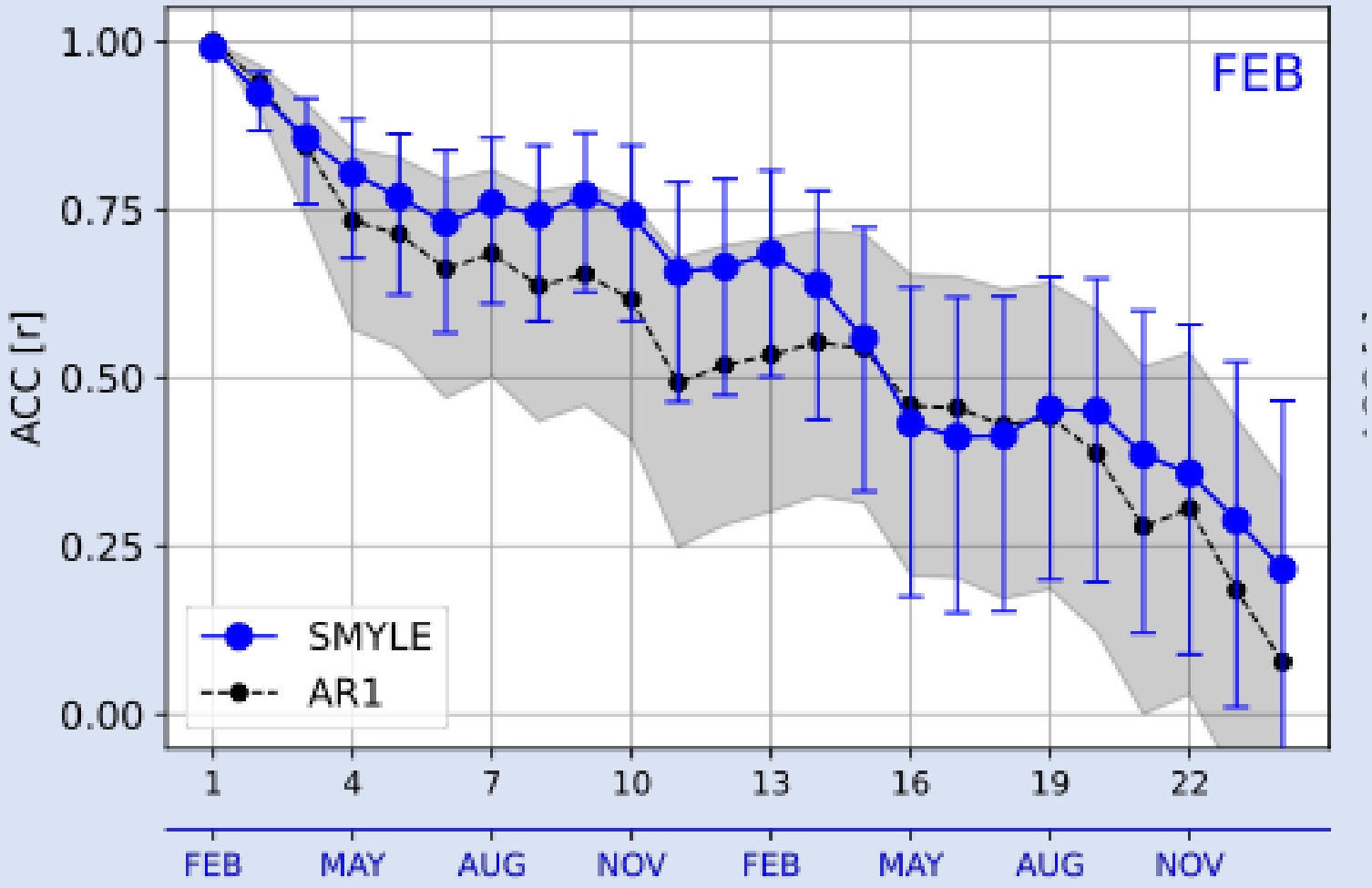


# SMYLE & AR1 uncertainty using standard confidence intervals



# SMYLE & AR1 uncertainty using standard confidence intervals

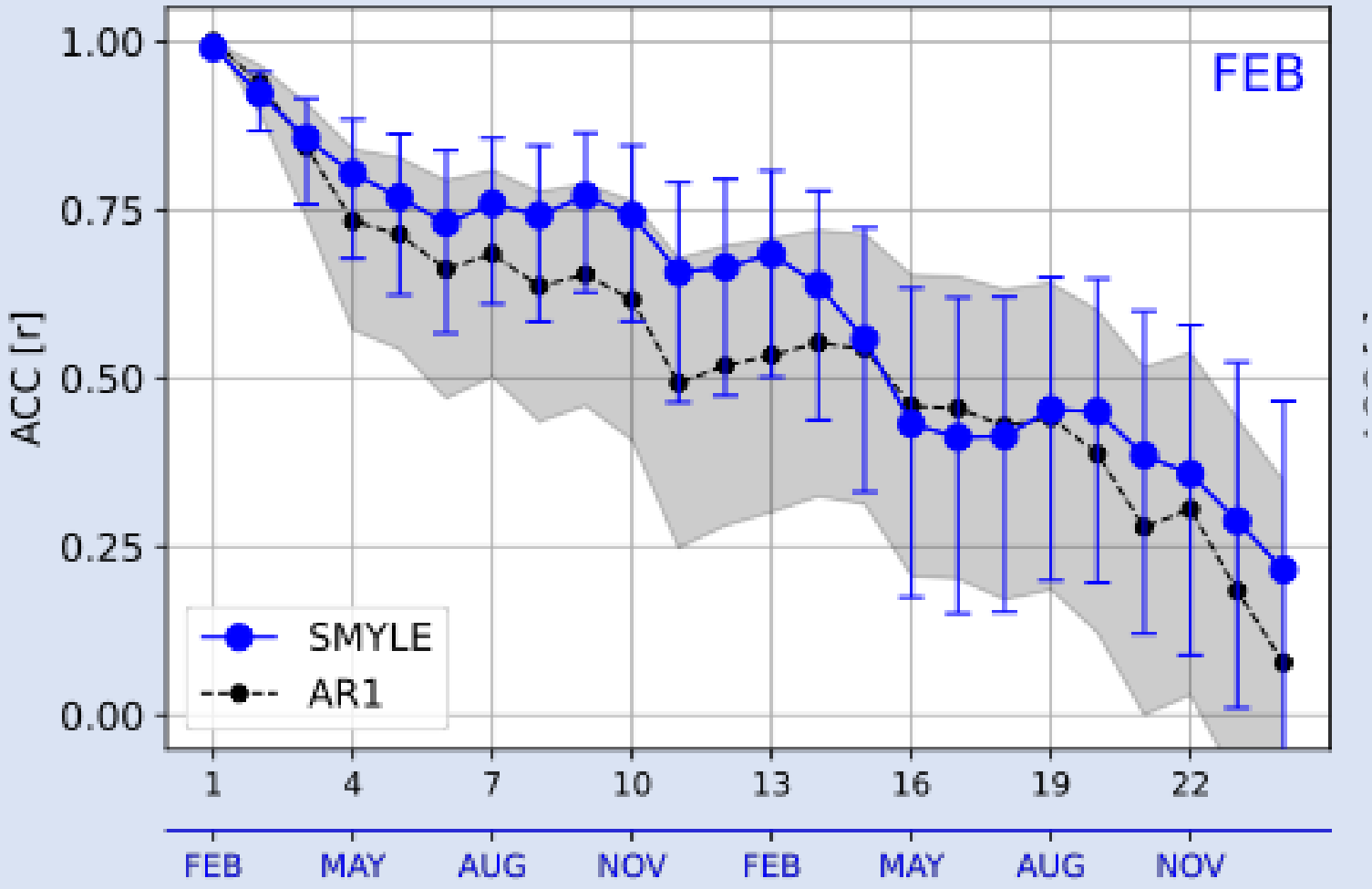
**Problem:** SMYLE and AR1 predictions are *not independent* because they are verified against the *same observations*.



# SMYLE & AR1 uncertainty using standard confidence intervals

**Problem:** SMYLE and AR1 predictions are *not independent* because they are verified against the *same observations*.

**Solution:** Use a significance test which *explicitly* accounts for *inter-model correlation*.



# Significance test explicitly accounting for inter-model correlation

$r_{ay}$: correlation between model A and observations

$r_{by}$: correlation between model B and observations

$r_{ab}$: correlation between model A and model B

$n$: number of samples

$$T_2 = (r_{by} - r_{ay}) \sqrt{\frac{(n-1)(1+r_{ab})}{2\left(\frac{n-1}{n-3}\right)R + \frac{1}{4}(r_{ay}+r_{by})^2(1-r_{ab})^3}} \qquad (7)$$

(Has a Student's *t* distribution with $n - 3$ degrees of freedom)

$$R = (1 - r_{ay}^2 - r_{by}^2 - r_{ab}^2) + (2r_{ay}r_{by}r_{ab}), \qquad (6)$$

(Determinant of 3x3 Correlation Matrix)

*Siegert et al., 2017*

# Significance test explicitly accounting for inter-model correlation

- Original: Hotelling (1940)
- Modified: Williams (1959)
- Suggested for psychological studies: Steiger (1980)
- Suggested for weather and climate prediction: Siegert et al. (2017)

$r_{ay}$: correlation between model A and observations
$r_{by}$: correlation between model B and observations
$r_{ab}$: correlation between model A and model B
$n$: number of samples

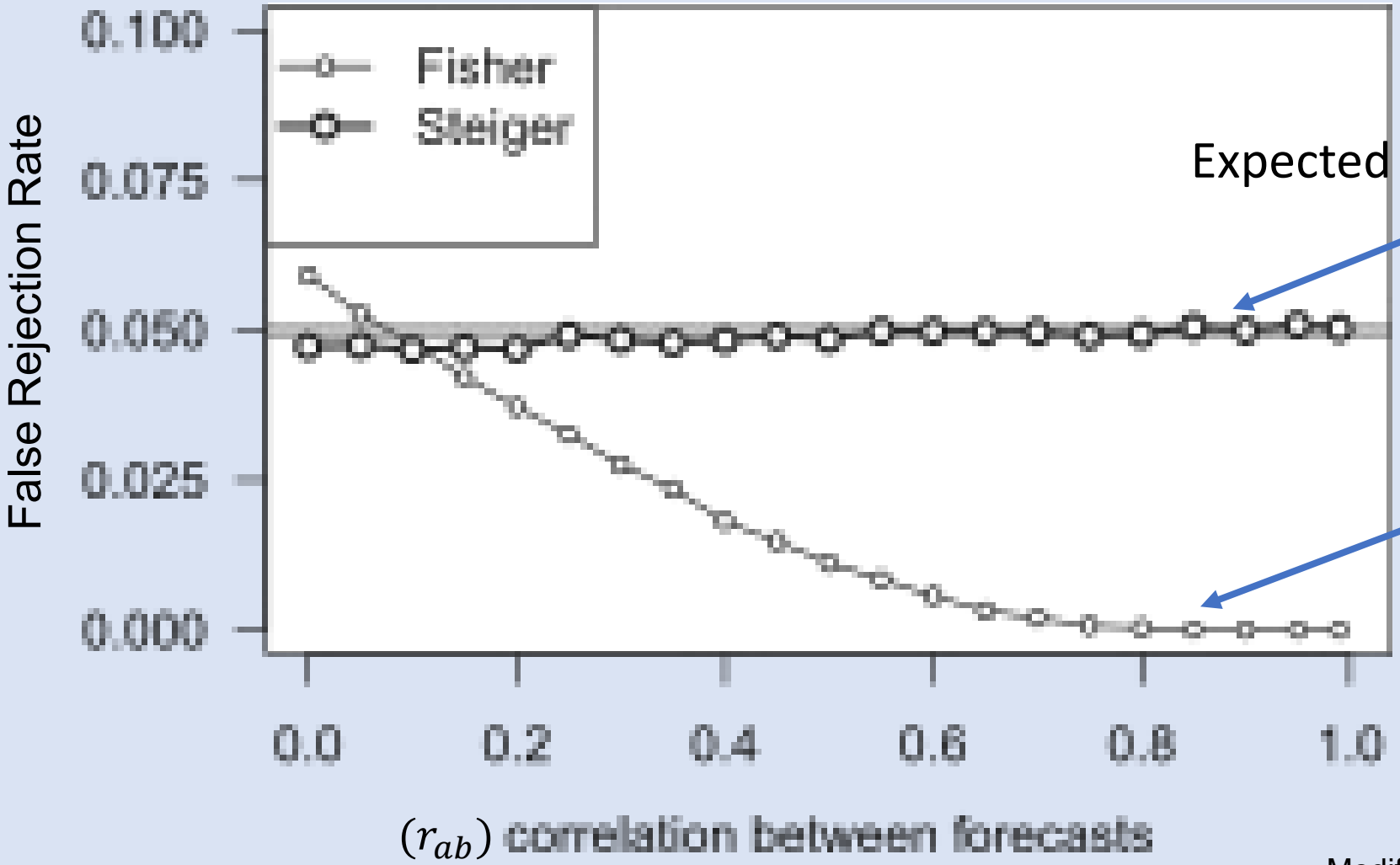
$$T_2 = (r_{by} - r_{ay}) \sqrt{\frac{(n-1)(1+r_{ab})}{2\left(\frac{n-1}{n-3}\right)R + \frac{1}{4}(r_{ay}+r_{by})^2(1-r_{ab})^3}} \qquad (7)$$

(Has a Student's $t$ distribution with $n-3$ degrees of freedom)

$$R = (1 - r_{ay}^2 - r_{by}^2 - r_{ab}^2) + (2r_{ay}r_{by}r_{ab}), \qquad (6)$$

(Determinant of 3x3 Correlation Matrix)

# “Steiger” test outperforms standard test for correlated models

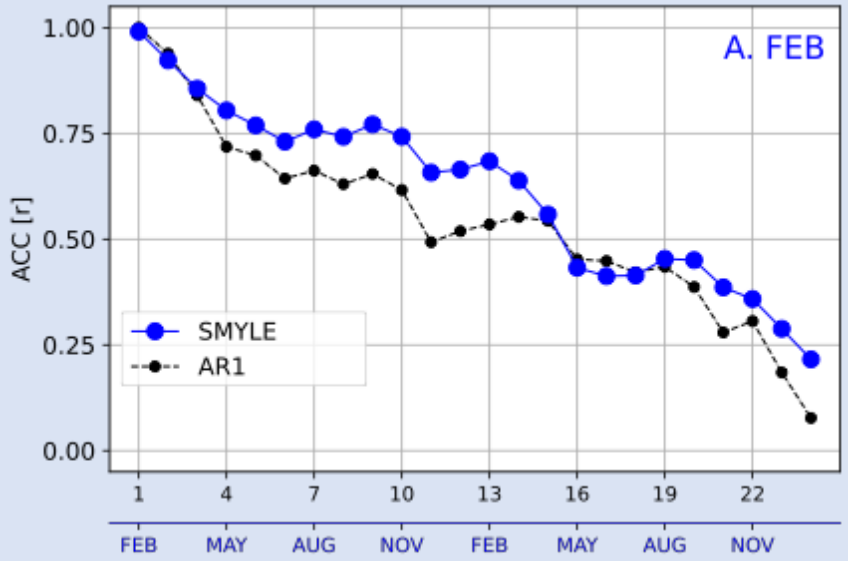



Expected false rejection rate

Too conservative

Modified from *Siegert et al. (2017)*

# Applying the test to SMYLE and AR1 PDO predictions




1. Calculate $r_{ay}$, $r_{by}$, and $r_{ab}$
2. Calculate $T_2$
3. Find $T_{crit}$ for your significance level and number of samples (e.g., $\alpha = 0.95,\ \ n = 50$)
4. If $T_2 > |T_{crit}|$, reject $H_0$

$r_{ay}$: correlation between model A and observations
$r_{by}$: correlation between model B and observations
$r_{ab}$: correlation between model A and model B
$n$: number of samples

$$T_2 = (r_{by} - r_{ay}) \sqrt{\frac{(n-1)(1+r_{ab})}{2\left(\frac{n-1}{n-3}\right)R + \frac{1}{4}(r_{ay}+r_{by})^2(1-r_{ab})^3}} \qquad (7)$$

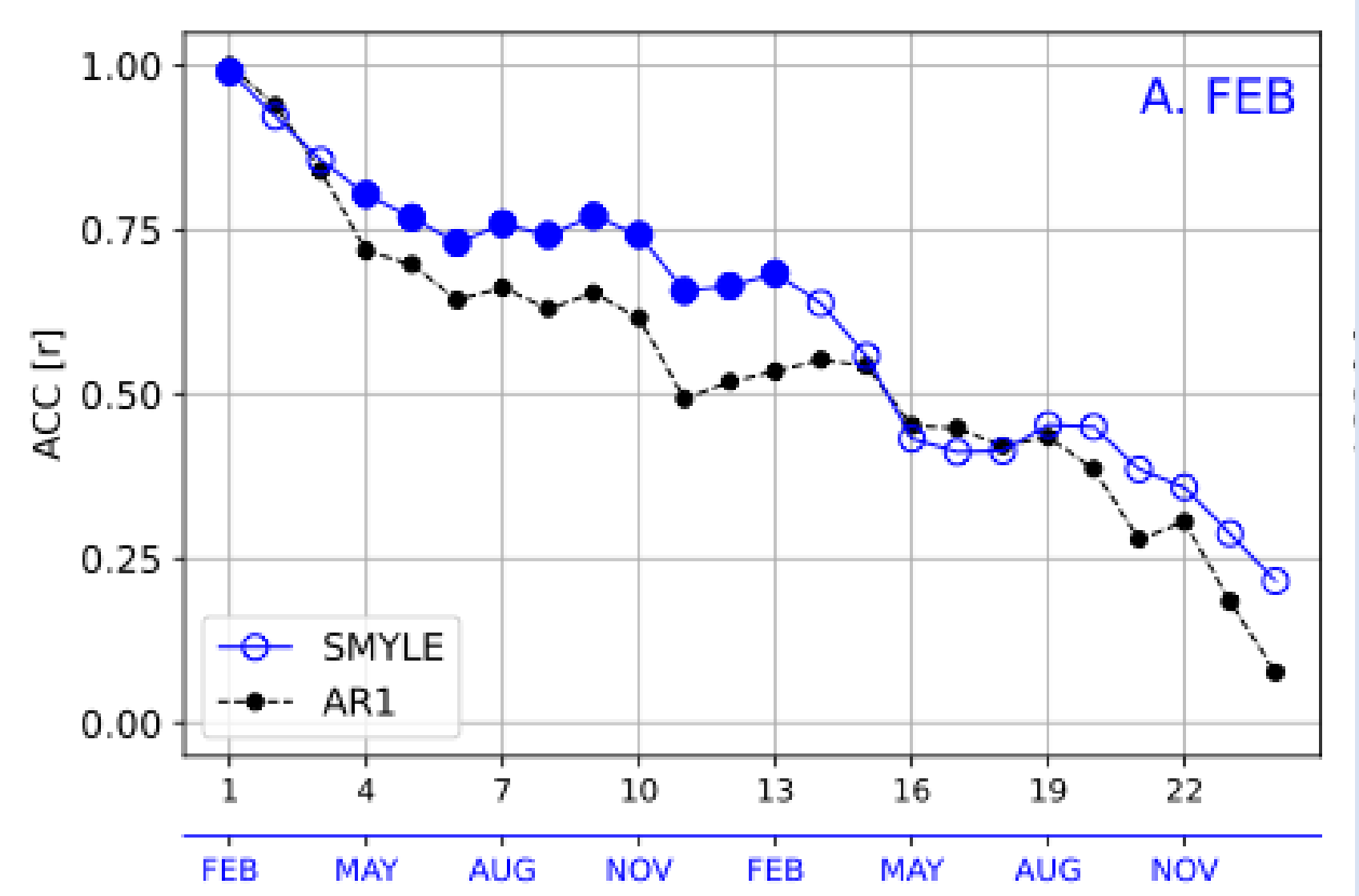
(Has a Student's $t$ distribution with $n-3$ degrees of freedom)

$$R = (1 - r_{ay}^2 - r_{by}^2 - r_{ab}^2) + (2r_{ay}r_{by}r_{ab}), \qquad (6)$$

(Determinant of 3x3 Correlation Matrix)

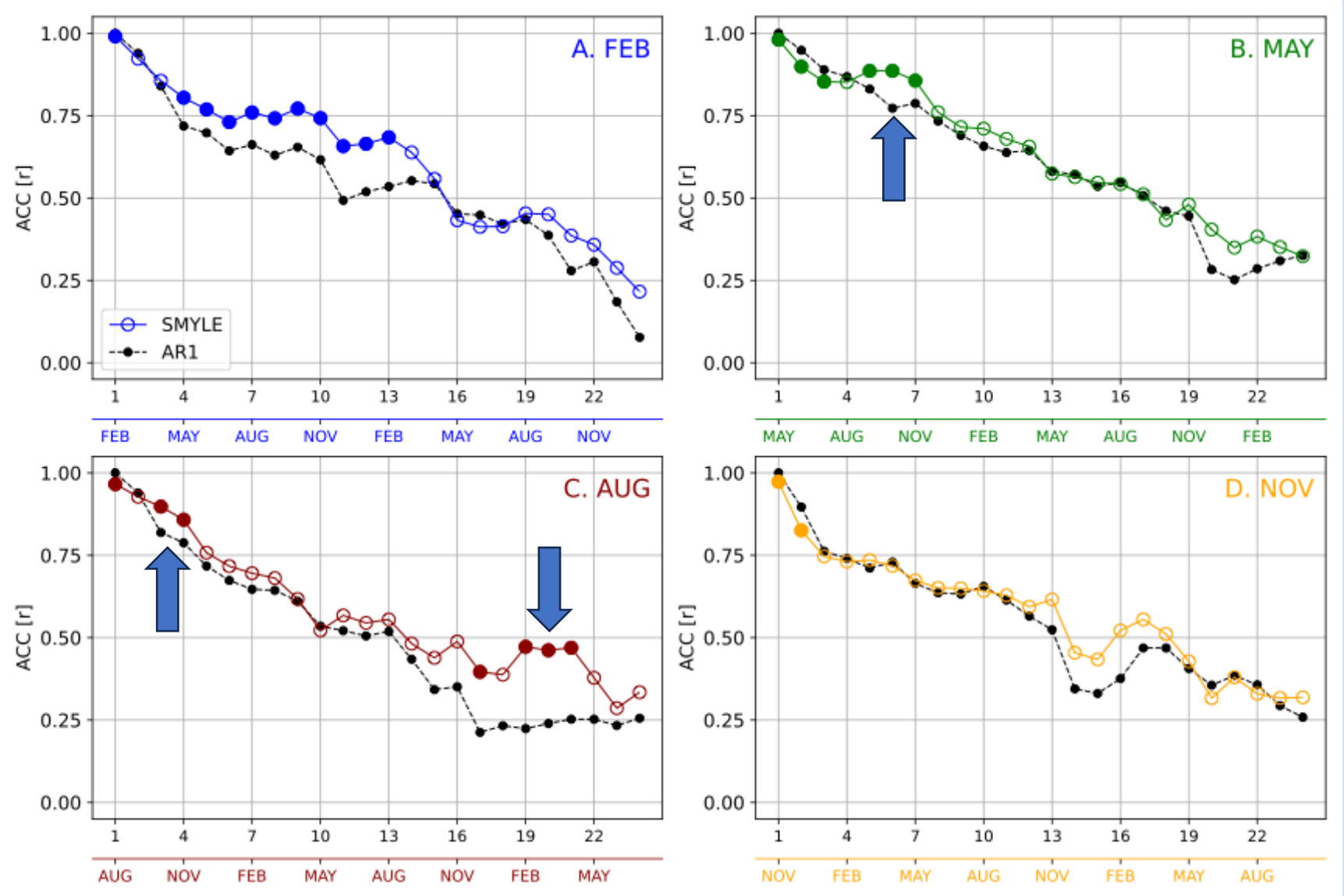
# SMYLE significantly outperforms AR1 from leads 4-13

- ● significant
- ○ not significant

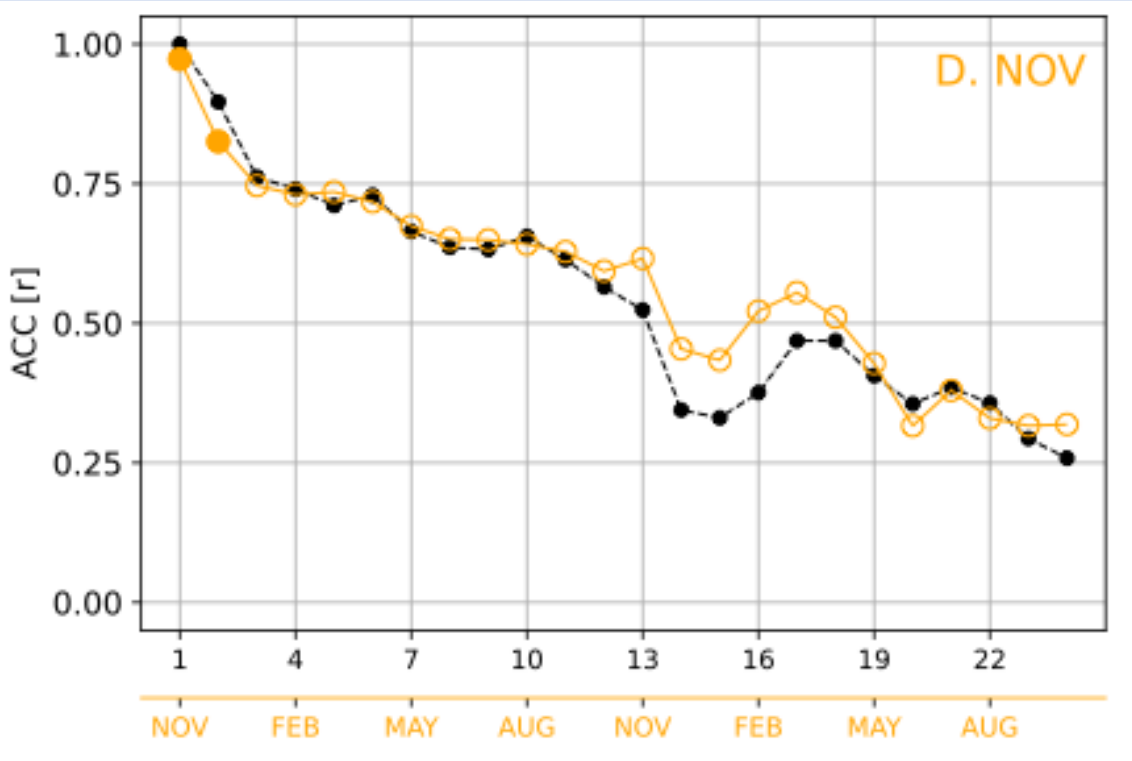


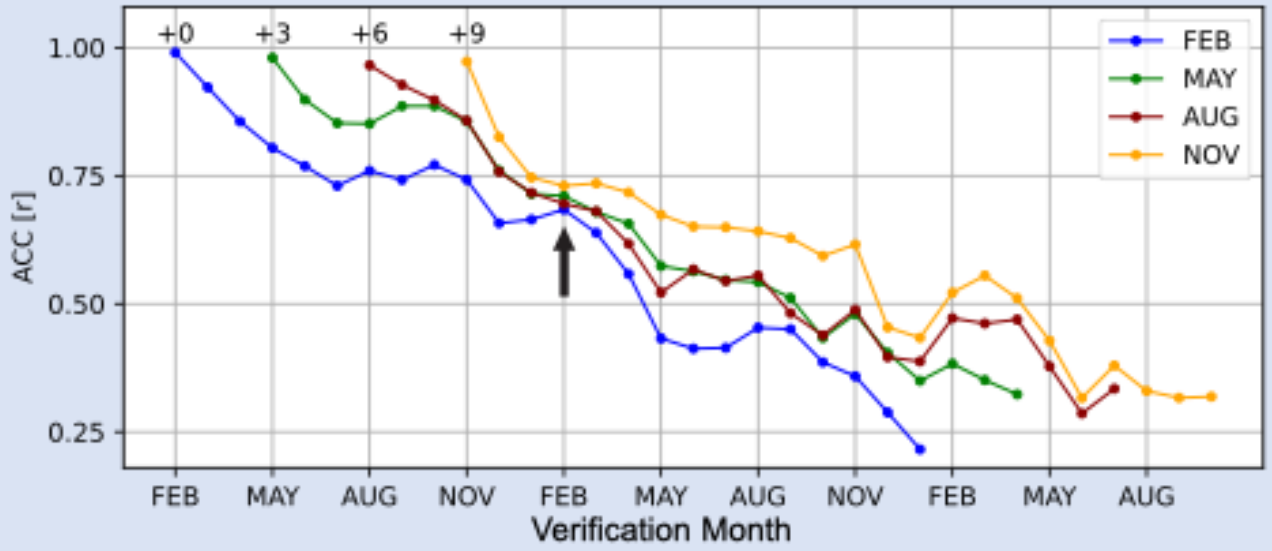
# SMYLE FEB has the longest lasting statistical significance
# SMYLE NOV is insignificant (or significantly *worse*) at all leads

- ● significant
- ○ not significant

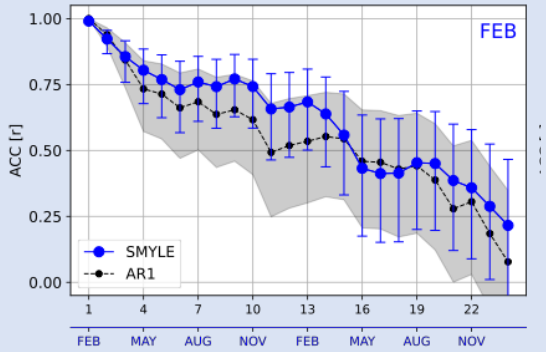


# SMYLE NOV struggles to predict the PDO

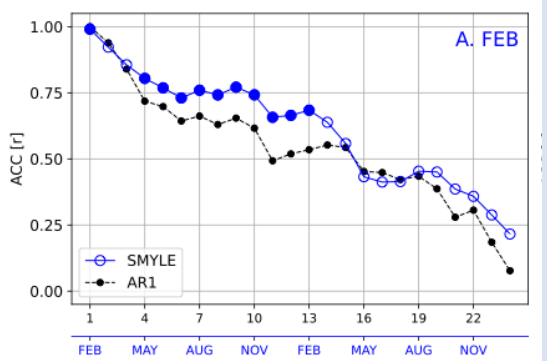


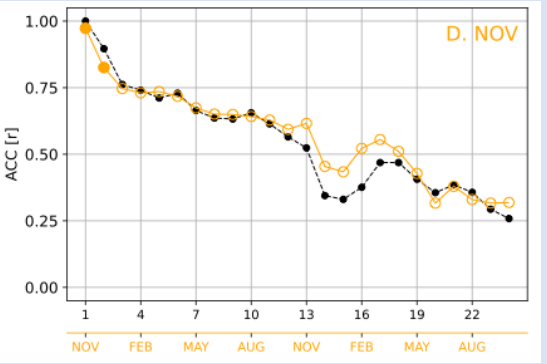
# Conclusions




*Inter-model correlation* must be accounted for when comparing model prediction skill. (see *Siegert et al., 2017*)




SMYLE PDO prediction skill *out-performs* AR1 predictions for **FEB** initialization from leads 4-13




**NOV** initialization does not significantly outperform AR1 at any lead time.

Full results to be published in GRL pending second review:
Meeker et al., *Seasonality of Pacific Decadal Oscillation Prediction Skill*

# References


Meeker, E., Maroon, E., Deppenmeier, A. L., Thompson, L., Vimont, D., & Yeager, S. G. (in revision). Seasonality of Pacific Decadal Oscillation Prediction Skill. *Geophysical Research Letters*, TBD.

---

Hotelling, H. (1940). The selection of variates for use in prediction with some comments on the general problem of nuisance parameters, *Ann. Math. Statist.,* 11, 271-283.

Siegert, S., Bellprat, O., Ménégoz, M., Stephenson, D. B., & Doblas-Reyes, F. J. (2017). Detecting Improvements in Forecast Correlation Skill: Statistical Testing and Power Analysis. Monthly Weather Review, 145(2), 437–450. https://doi.org/10.1175/MWR-D-16-0037.1

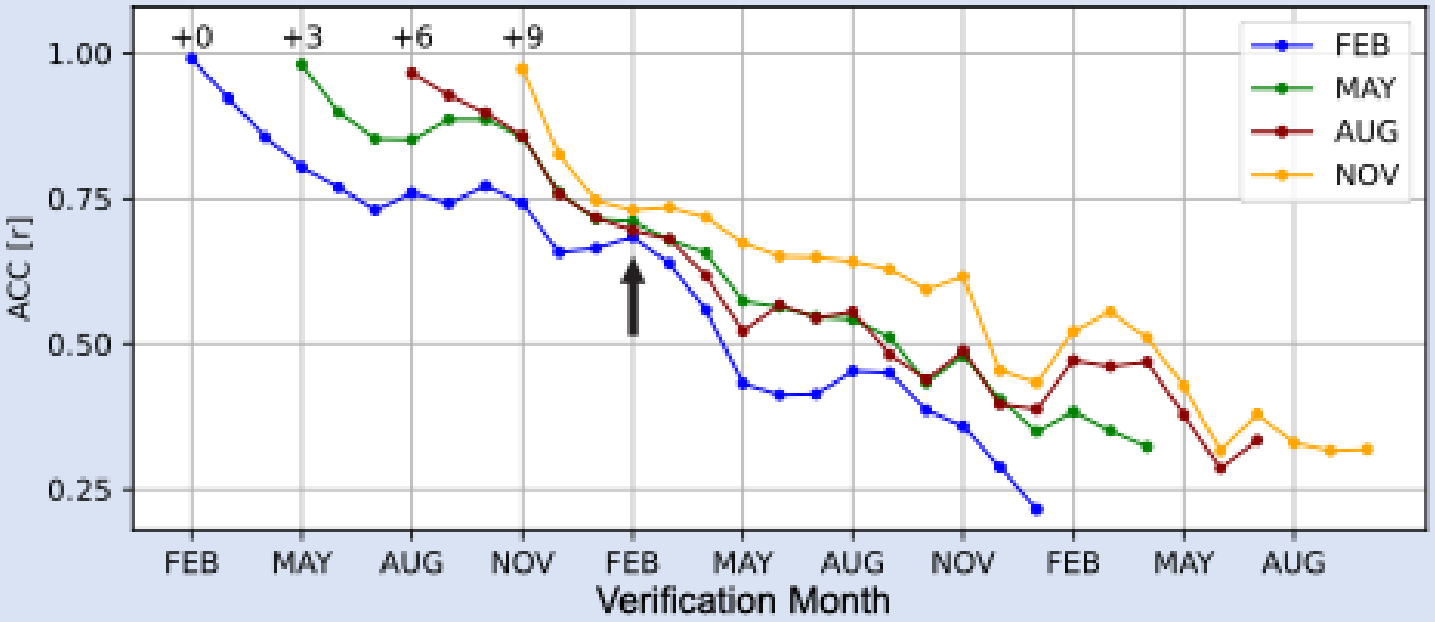
Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. Psychological Bulletin, 87(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

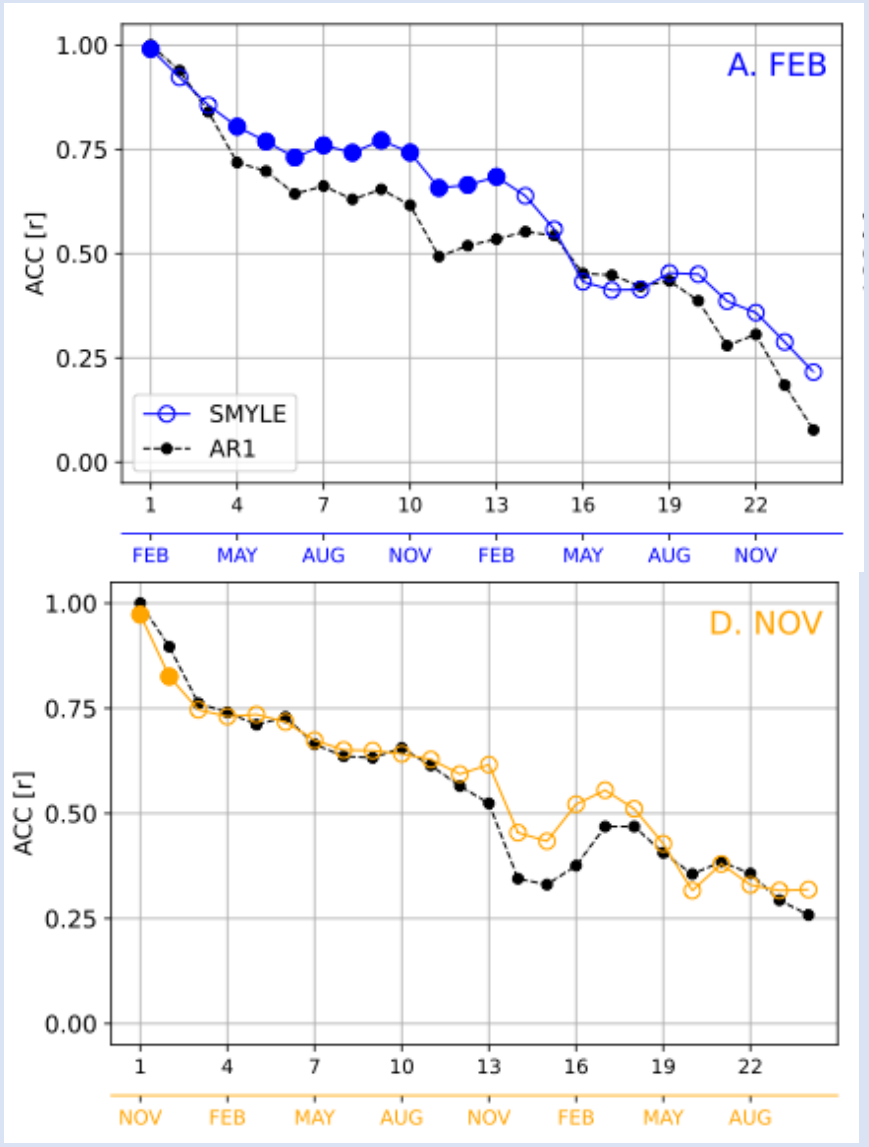
Williams, E. J. (1959). The comparison of regression variables. *J. Roy. Stat. Soc. B*, 21 (2), 396–399. http://www.jstor.org/stable/2983809.

Yeager, S. G., Rosenbloom, N., Glanville, A. A., Wu, X., Simpson, I., Li, H., Molina, M. J., Krumhardt, K., Mogen, S., Lindsay, K., Lombardozzi, D., Wieder, W., Kim, W. M., Richter, J. H., Long, M., Danabasoglu, G., Bailey, D., Holland, M., Lovenduski, N., … King, T. (2022). The Seasonal-to-Multiyear Large Ensemble (SMYLE) prediction system using the Community Earth System Model version 2. *Geoscientific Model Development*, *15*(16), 6451–6493. https://doi.org/10.5194/gmd-15-6451-2022

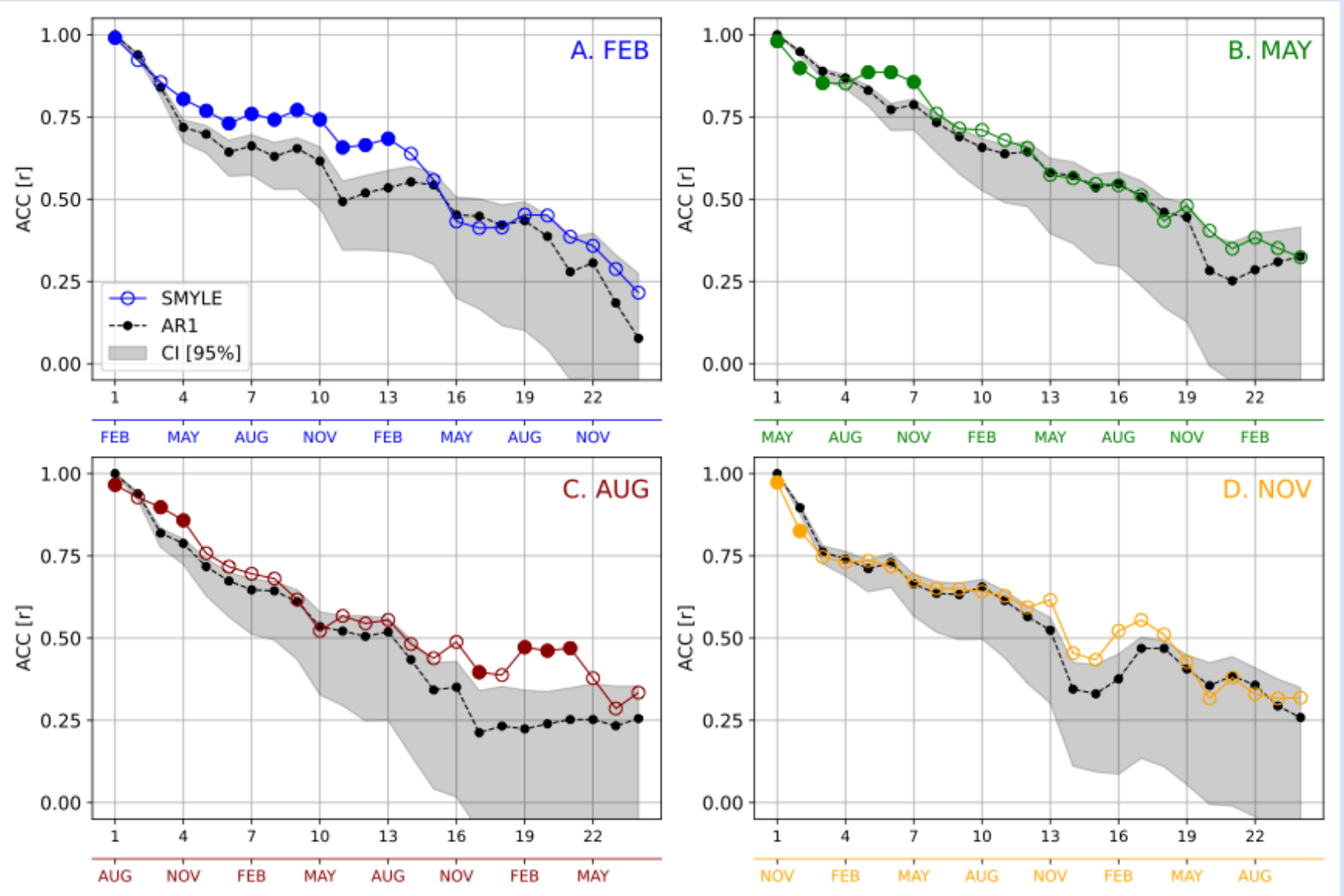
# Questions?

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON
emeeker2@wisc.edu



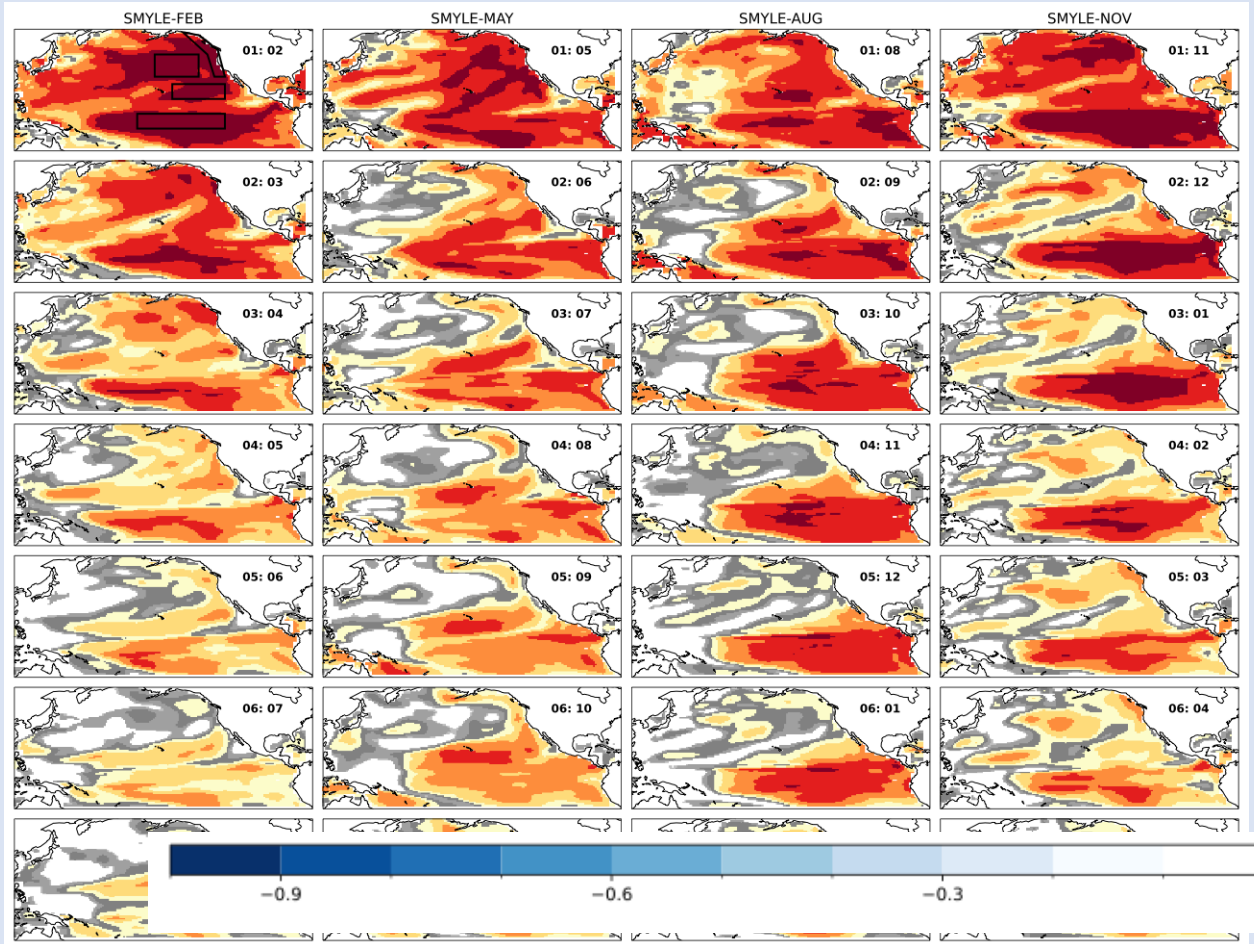



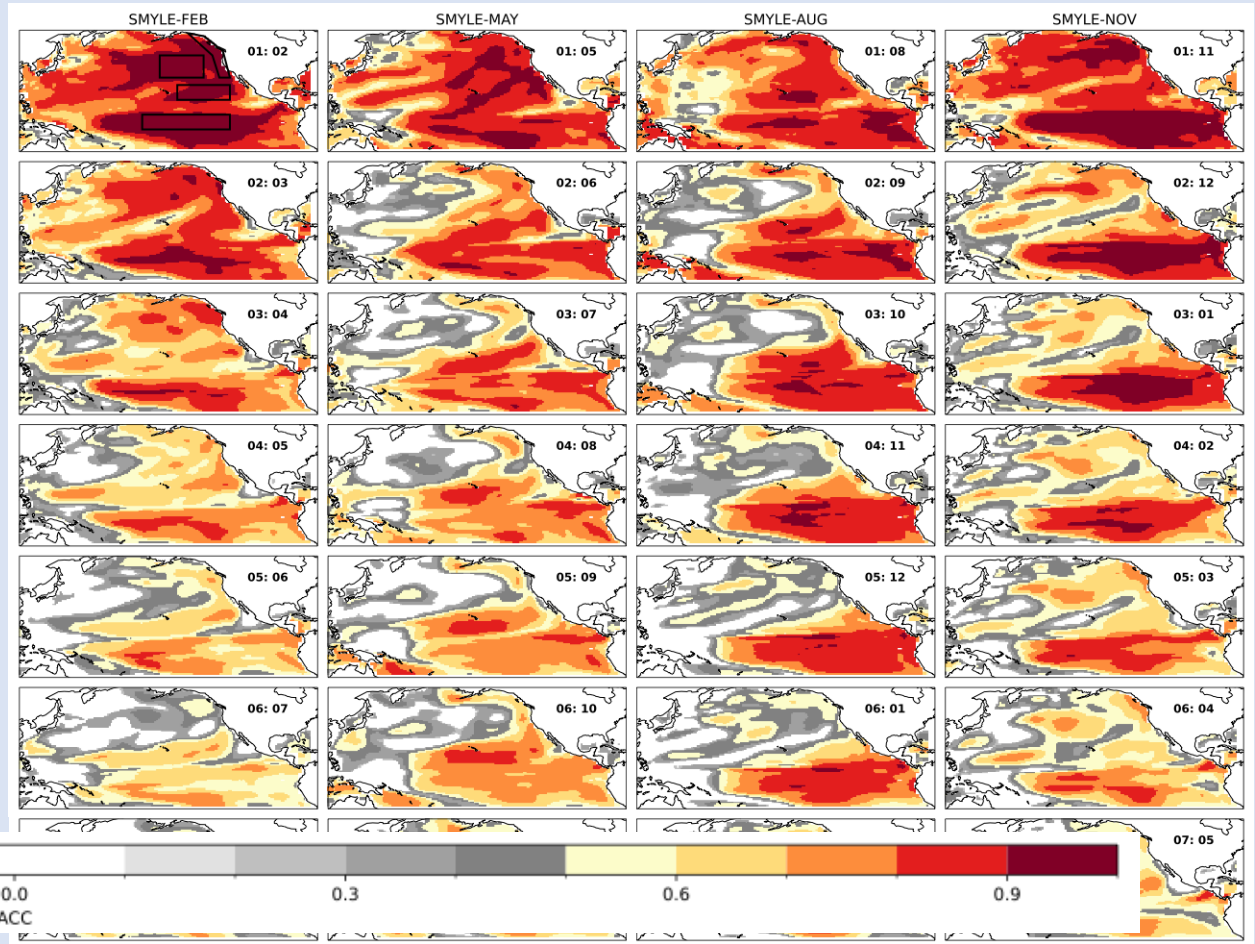
# Extra Slides



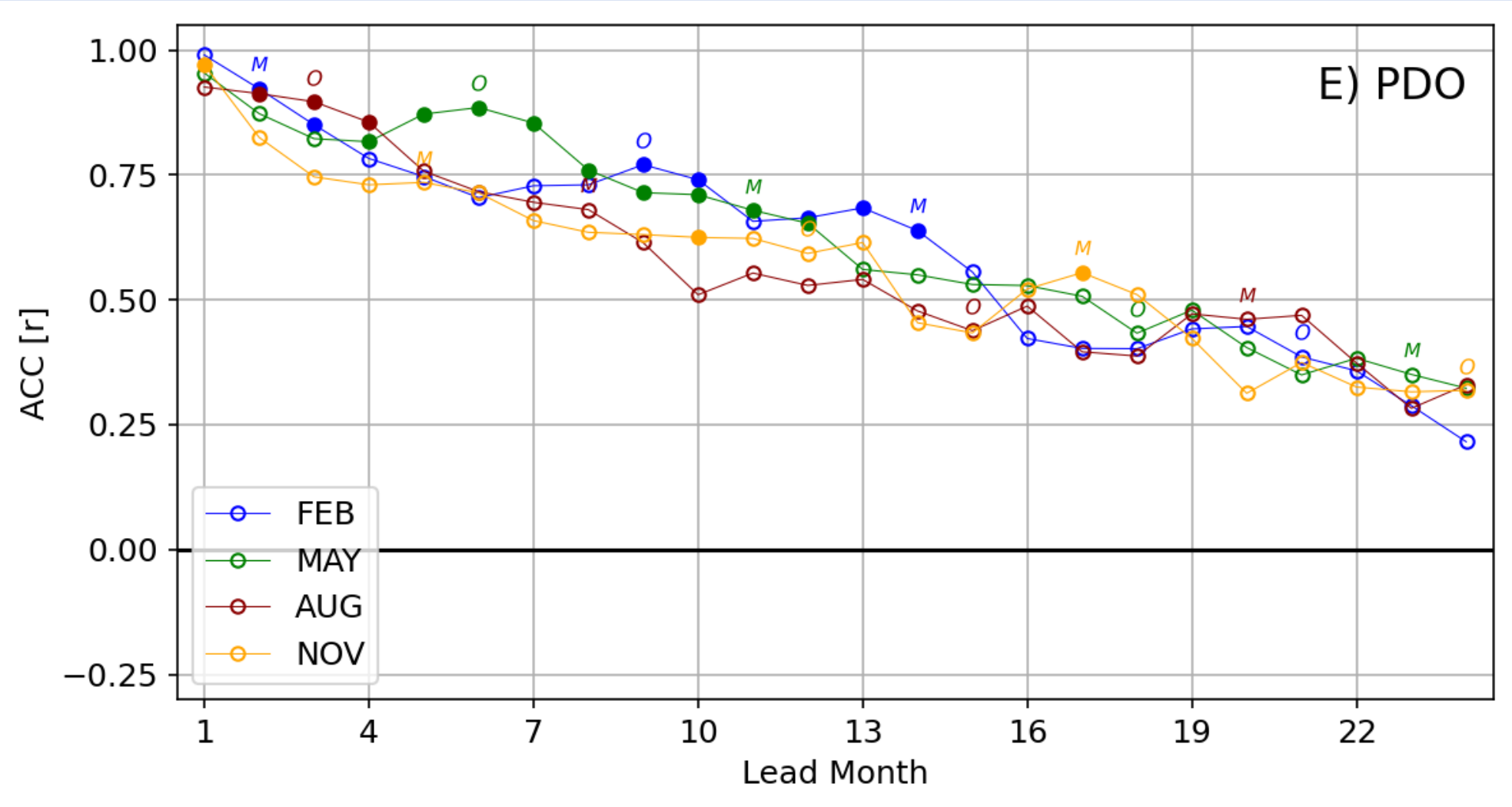

$$P_t = \alpha P_{t-1} + \sigma \eta_t$$

# Extra Slides

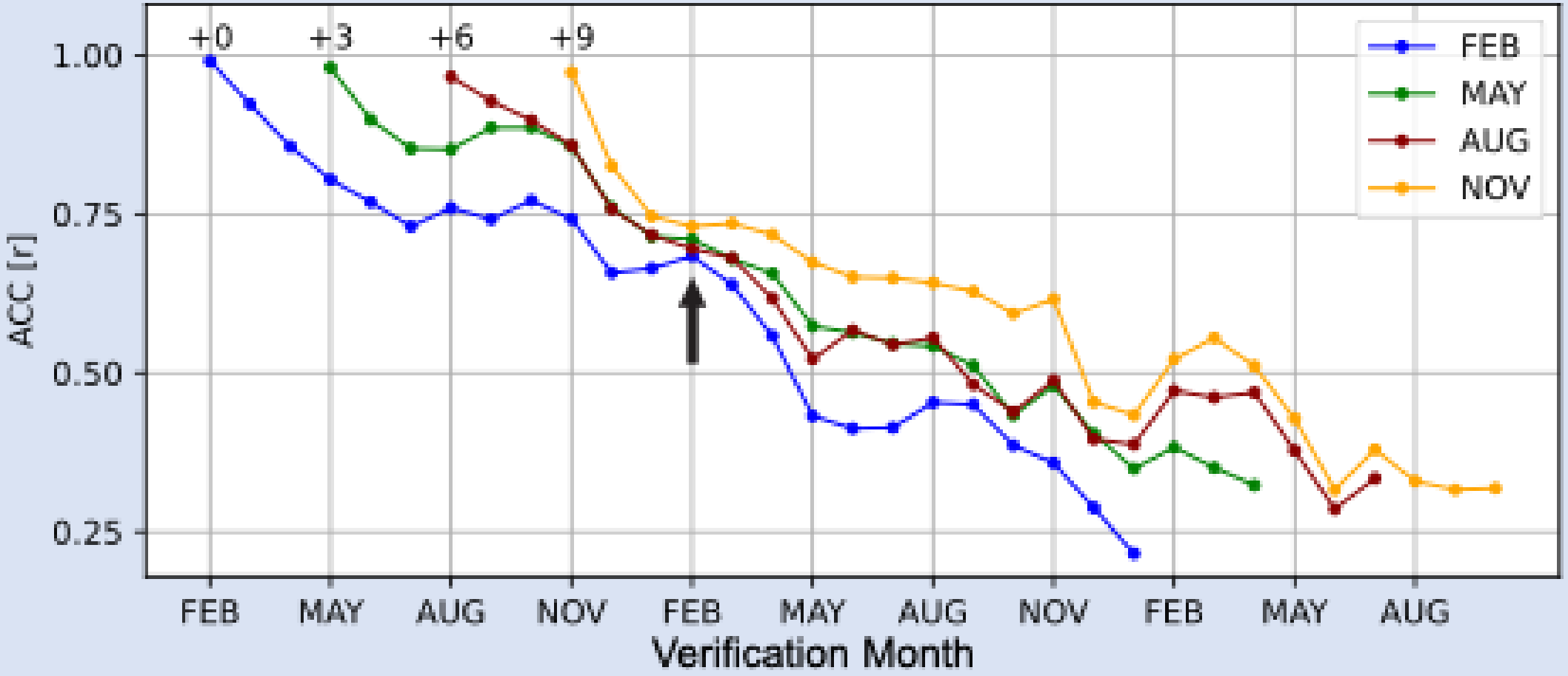


# Extra Slides



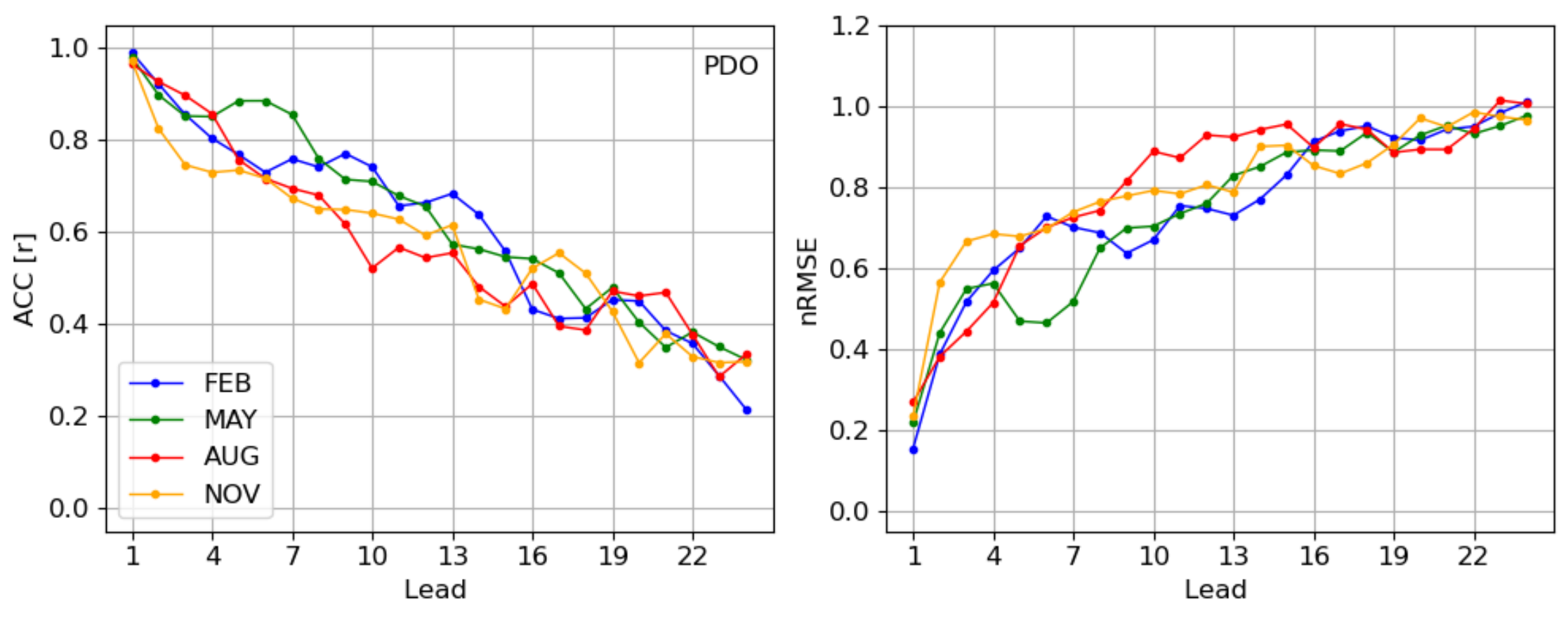
# Extra Slides



# Extra Slides

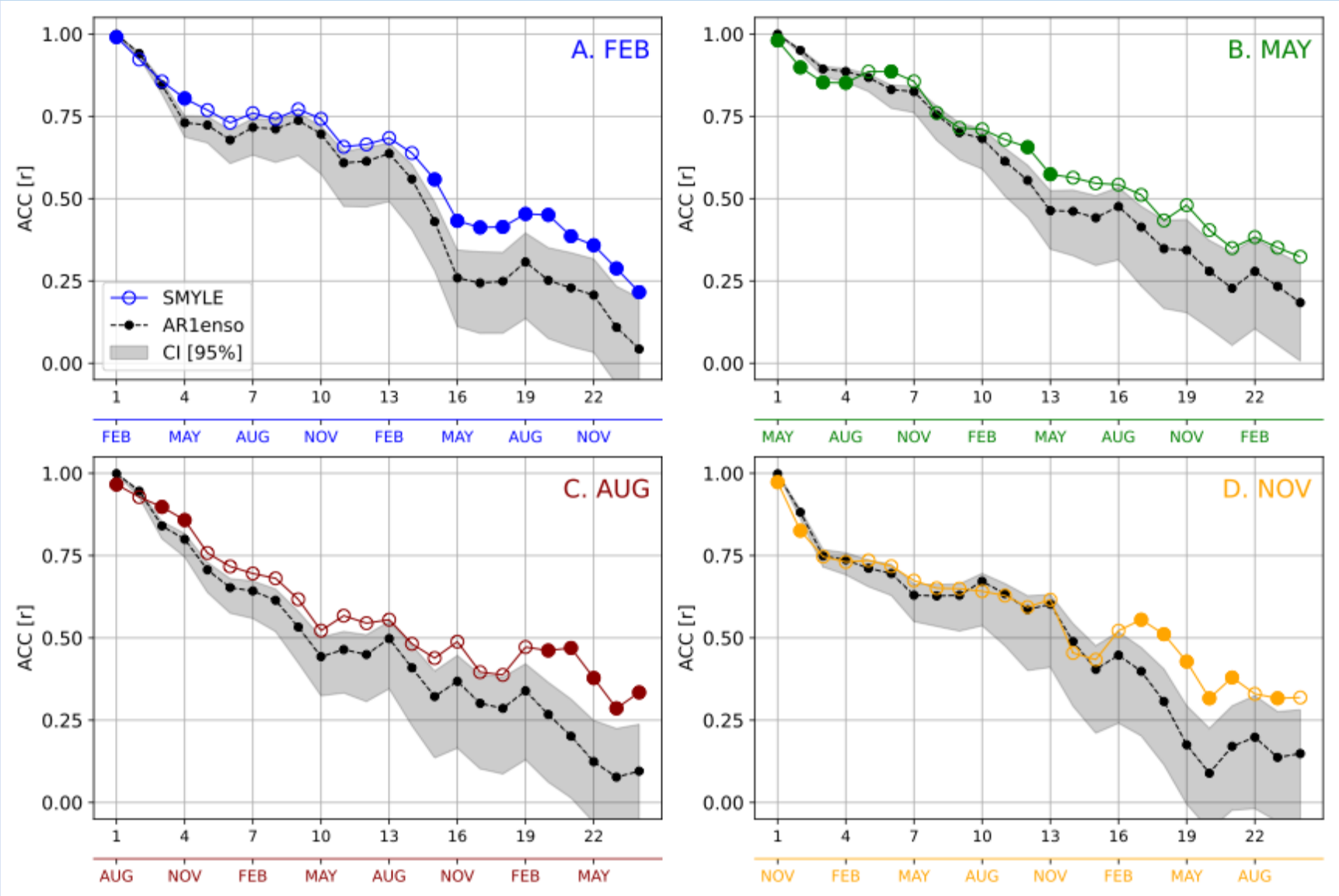
ACC

(normalized RMSE)



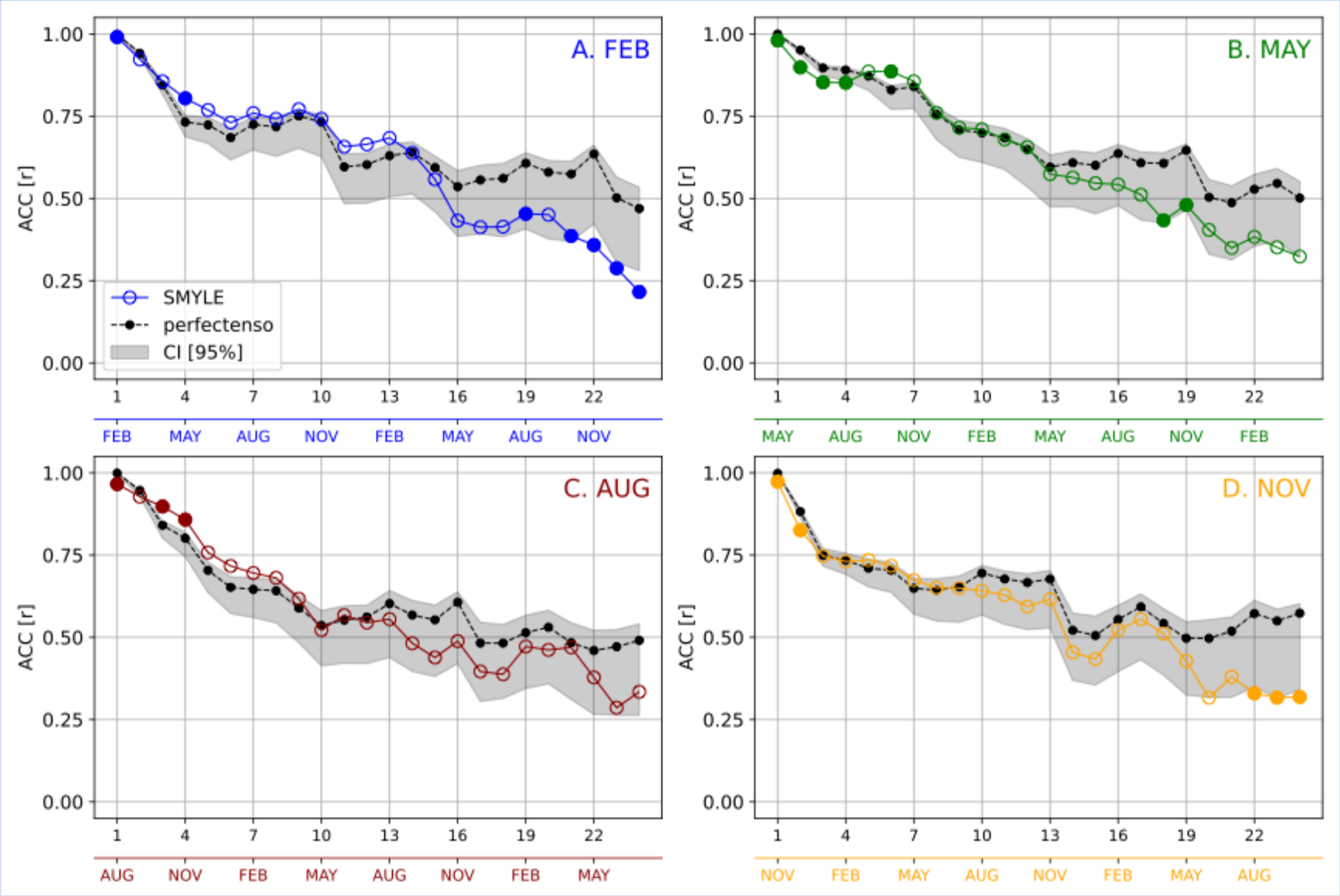
# Extra Slides

AR1+ENSO



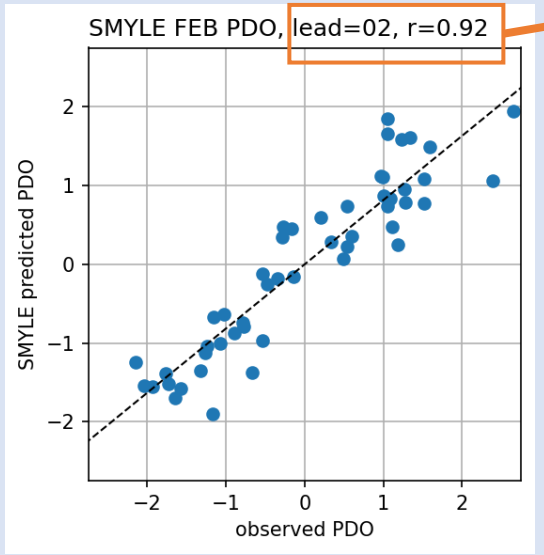
# Extra Slides

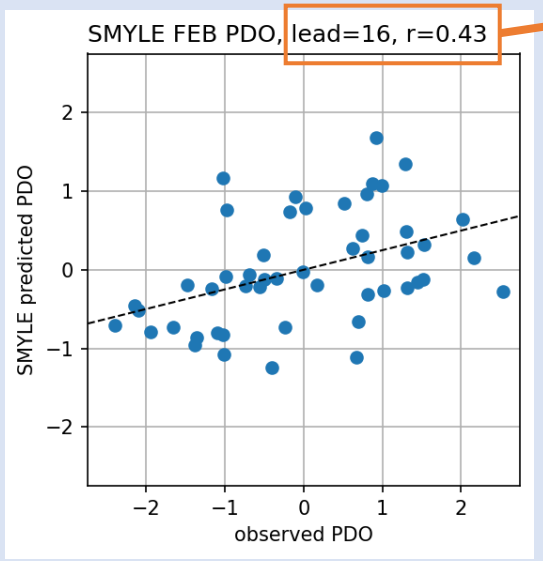
Perfect-ENSO

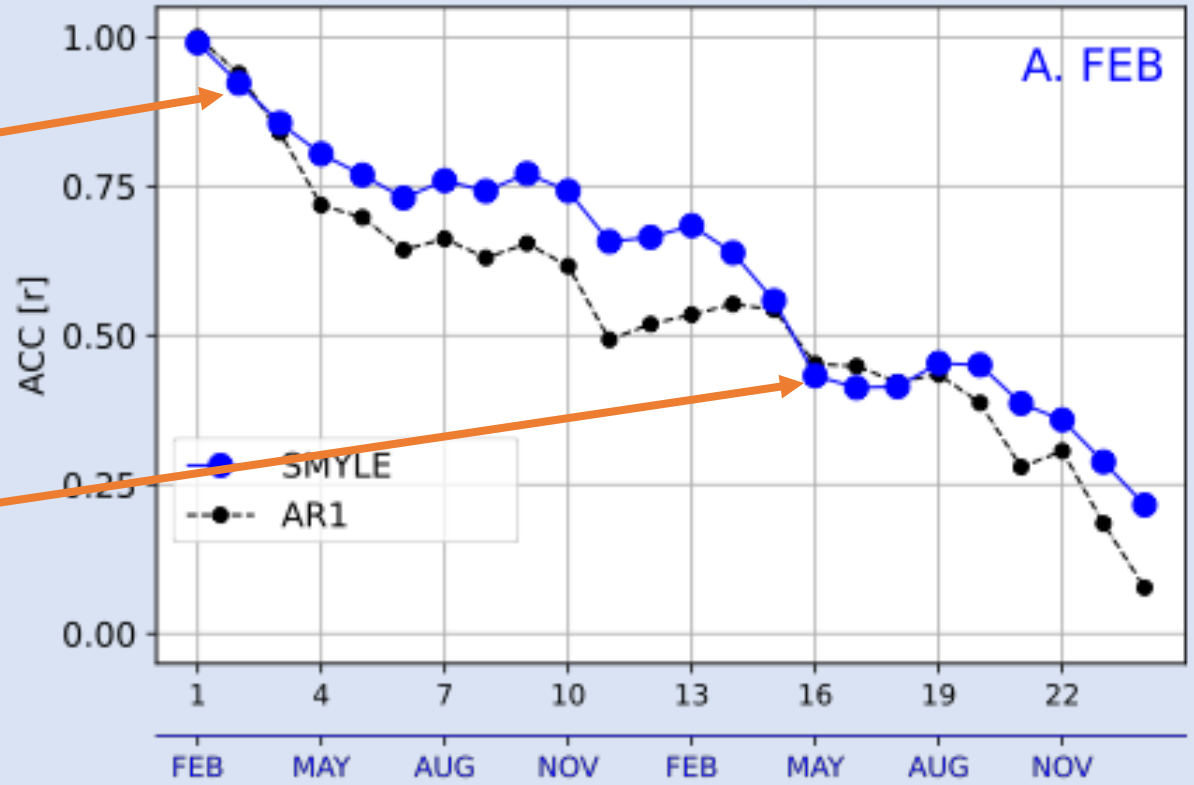


# Bookkeeping: ACC & AR1

- Anomaly Correlation Coefficient (ACC)
  - How well do predictions for a given lead correlate with observations?
  - Doesn't account for magnitude of error

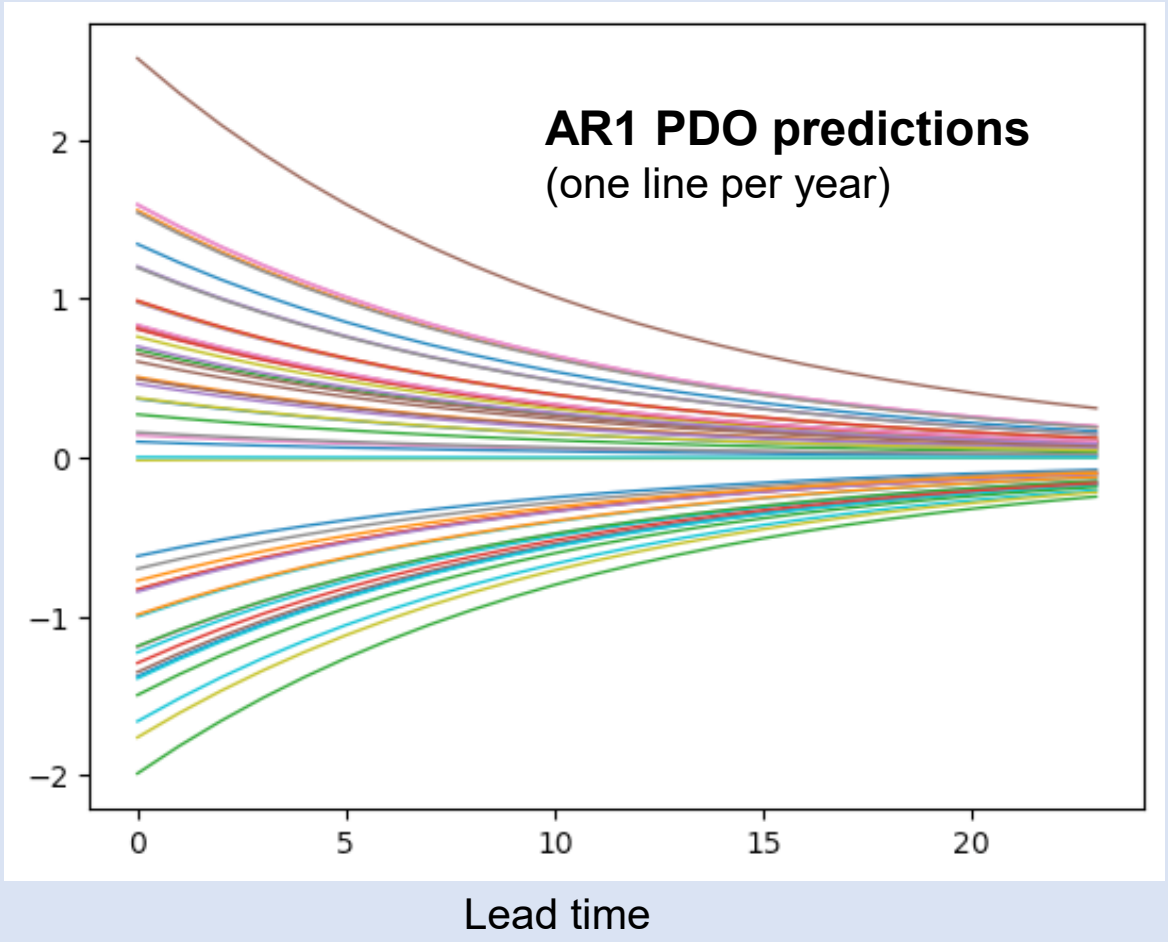


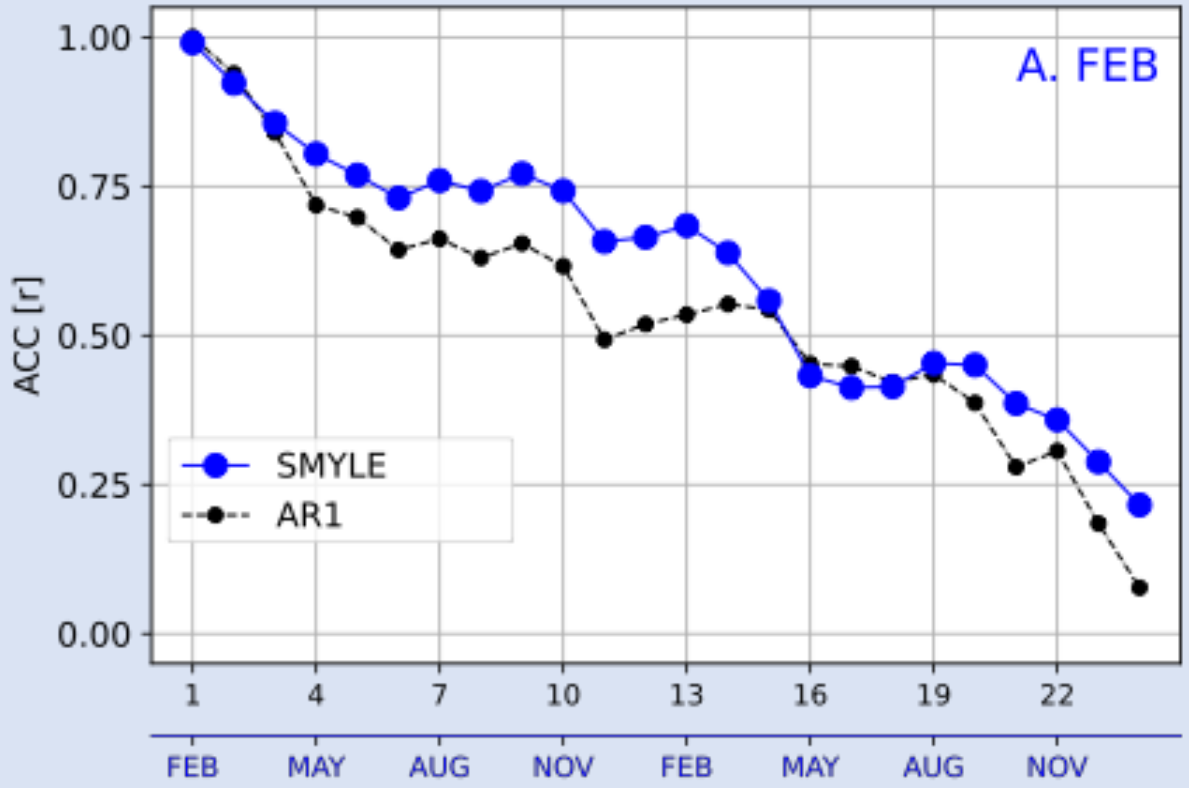



# Bookkeeping: ACC & AR1

$$P_t = \alpha P_{t-1} + \sigma \eta_t$$




$$\alpha_{PDO} = 0.91$$