

# EARLY PREDICTION OF WEAK STRATOSPHERIC POLAR VORTEX STATES USING A CAUSAL PRECURSOR DETECTION SCHEME

Authors: Marlene Kretschmer<sup>1,2</sup>, Jakob Runge<sup>3</sup>, Dim Coumou<sup>1,4</sup>

<sup>1</sup>Potsdam Institute for Climate Impact Research, Germany, <sup>2</sup>Potsdam University, Germany, <sup>3</sup>Imperial College London, Great Britain, <sup>4</sup>VU University Amsterdam, Netherlands

## INTRODUCTION

Weak stratospheric polar vortex (SPV) states can cause extreme winter weather in the Northern Hemisphere mid-latitudes. Including stratosphere activity in seasonal model predictions therefore significantly improves skill. Though dynamical models have useful prediction skill of the SPV on synoptic time-scales of about one week, they poorly capture low-frequency processes and therefore have limited or no skill at longer lead-times [1]. Here we introduce a novel empirical prediction approach which builds upon response-guided community detection [2] and a causal discovery algorithm [3, 4].

## DATA

We study half-monthly data (ERA-Interim) of linearly detrended climatological anomalies from 1979 – 2015.

SPV Index		Precursor Fields
<b>Variable</b>	GPH at 10mb	SLP, SSTs, GPH at 500mb (Z500), poleward heat flux at 100mb (vT100), zonal wind at 50mb (u50)
<b>Location</b>	Polar cap mean, northward of 60° N	Gridded data, 20° S to 89° N
<b>Time</b>	November-March	Lead-lags up to four months (lag-8)

Tab.1: Table of considered data for the response variable and the precursors.

## METHOD

We calculate correlation maps of the precursor fields with the SPV index for different lags (first panel in Fig.1). Grid-points with significant correlation indices of same sign are grouped to communities and their area-weighted mean is computed to create a time-series for each region (second panel in Fig.1).

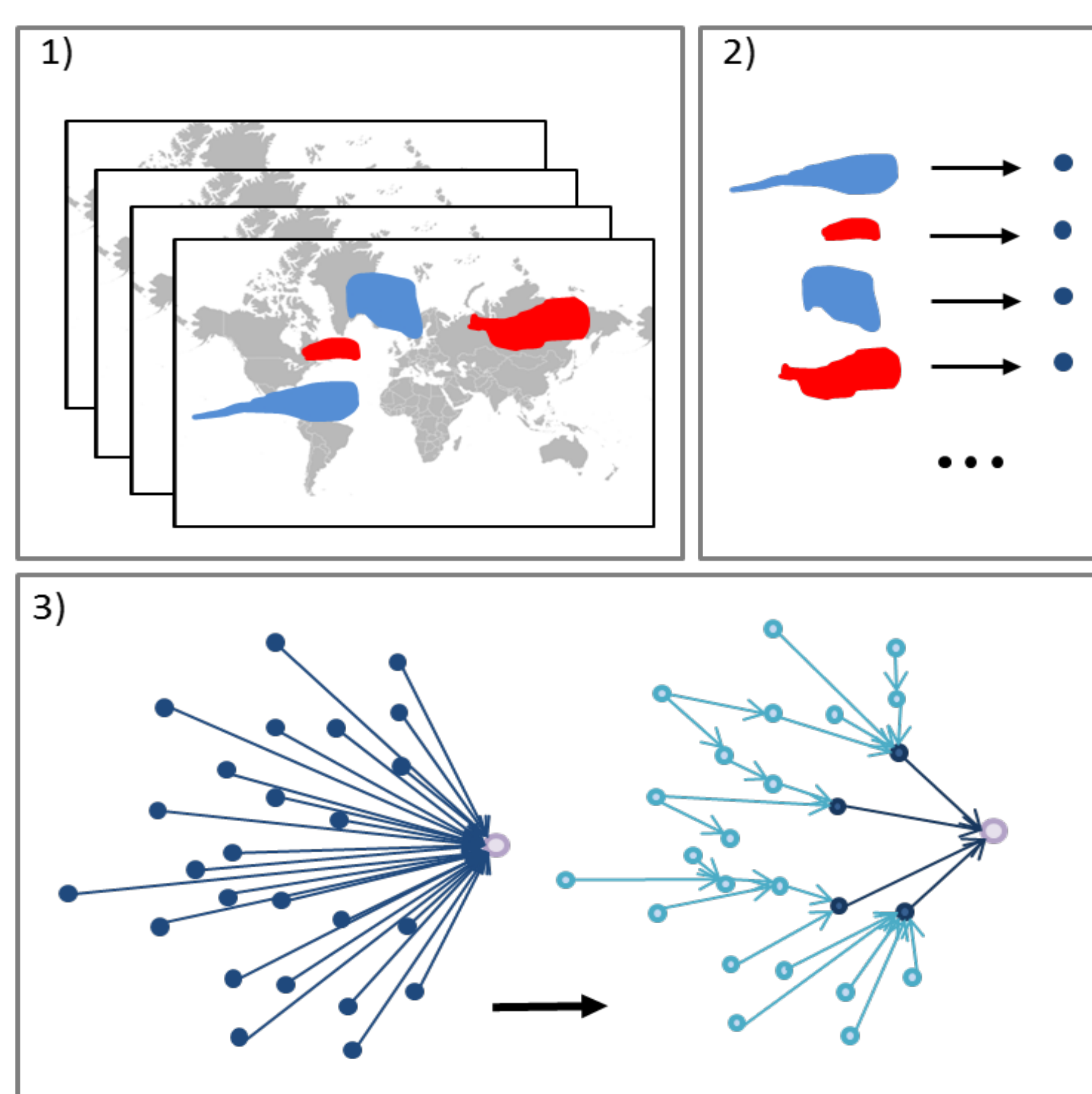


Figure 1: Schematic representation of the response-guided Causal Precursor Detection (CPD) scheme.

We then use a multivariate causal discovery algorithm to identify spurious correlations with the SPV due to auto-correlation, common drivers or indirect links (third panel in Fig.1). The remaining links (right part of third panel in Fig.1) are then interpreted as the causal drivers of the SPV.

## RESULTS & DISCUSSION

- We identify 471 precursor communities composed of all fields: vT100 (72), SLP (40), SST (278), Z500 (49) and u50 (32)
- Only 4 causal precursors remain after applying the causal discovery algorithm:
  - the lag-1 auto-correlated SPV index
  - two lag-1 vT100 regions over Eurasia and the North Pacific
  - a lag-3 community of SSTs in the tropical Philippine Sea
- The identified causal precursors are physically interpretable and mostly robust for different parameter settings

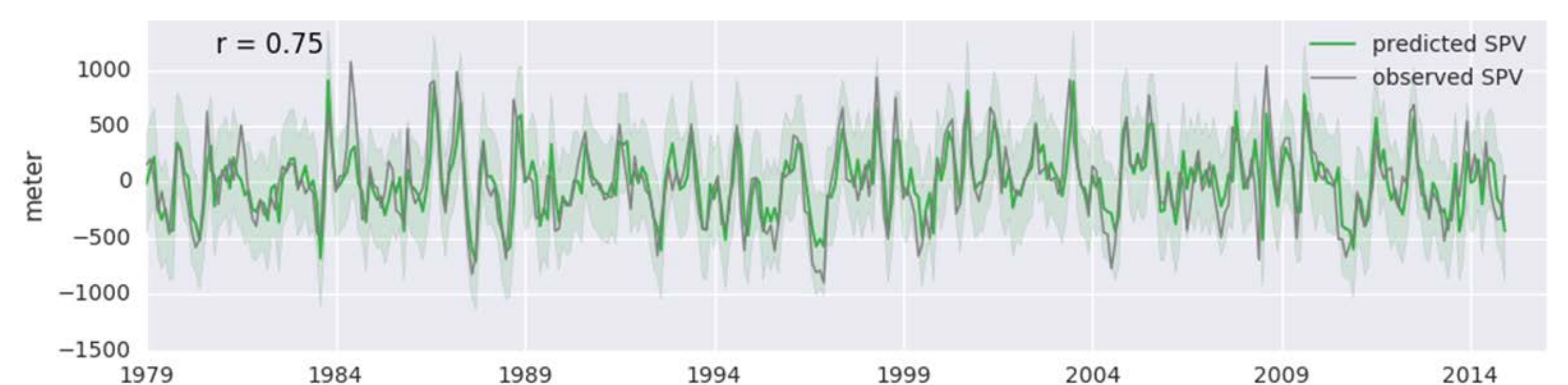


Figure 2: Cross validation of the regression model: One-step ahead predictions with the causal precursors as regressors. The regression models were built iteratively by leaving out data of the to predicted winter.

- A cross-validated multiple linear regression model of the causal precursors on the SPV can explain most of SPV variability ( $r^2=0.6$ )
- 64% of weak SPV states are accurately predicted with a false-alarm-rate of only  $\sim 4\%$  (odds-ratio = 42.3)
- For longer lead-lags 42% (16-30 days ahead), 22% (31-45 days ahead) and 14% (46-60 days ahead) of weak SPV events are still correctly predicted

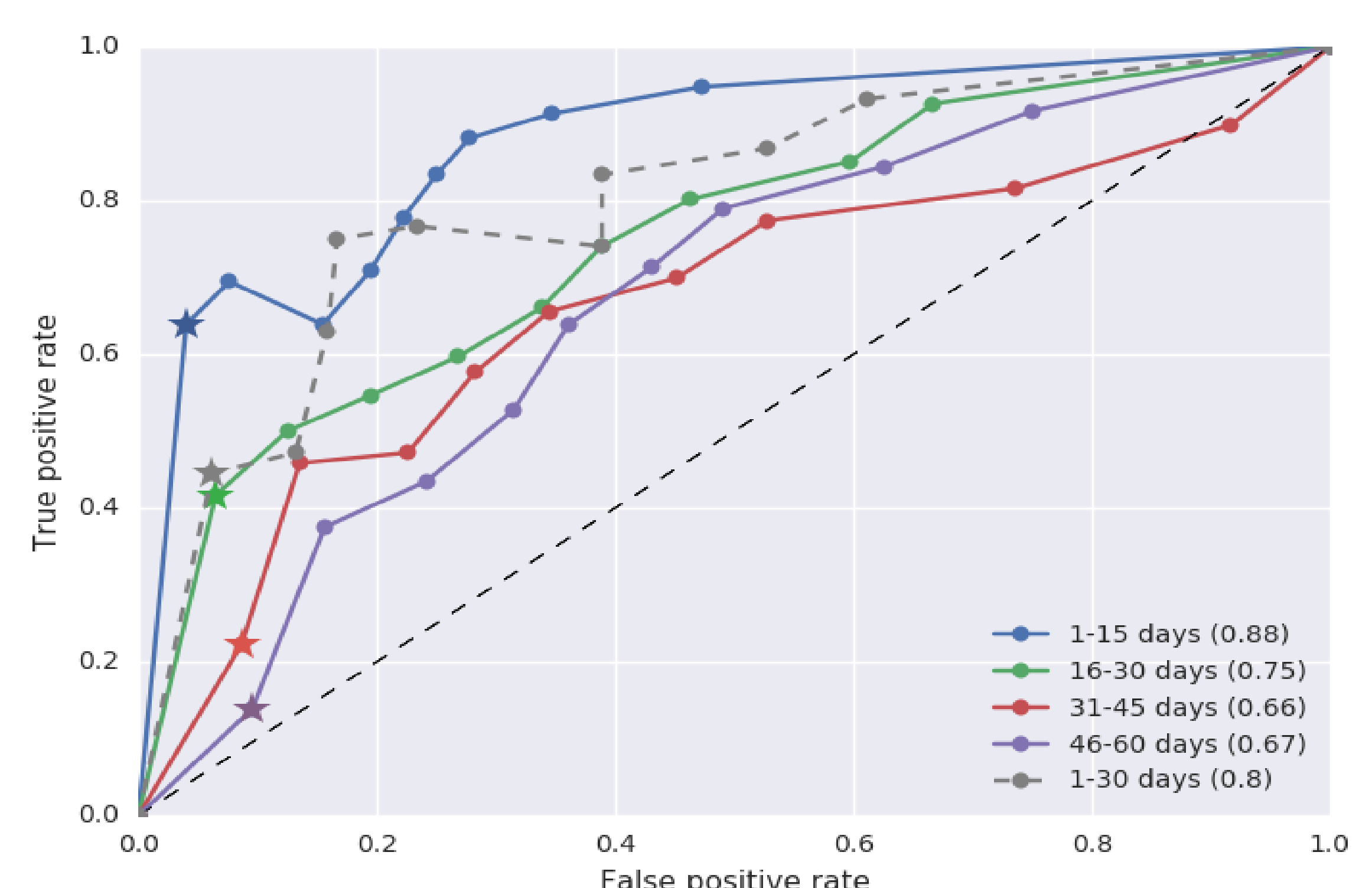


Figure 3: ROC scores for predictions based on different minimum lead-lags.

## CONCLUSION

The presented data-driven algorithm proposes a way to objectively identify linear (causal) precursors of a variable of interest. It gives robust and physically interpretable results for the SPV with high prediction skills for lead-times up to two months. Our approach can be used for early warning directly but also to select individual members from ensembles of operational forecast models to improve (sub-)seasonal winter predictions. The approach is generic in nature and can be applied to other climate indices to predict and understand causal links of teleconnections.