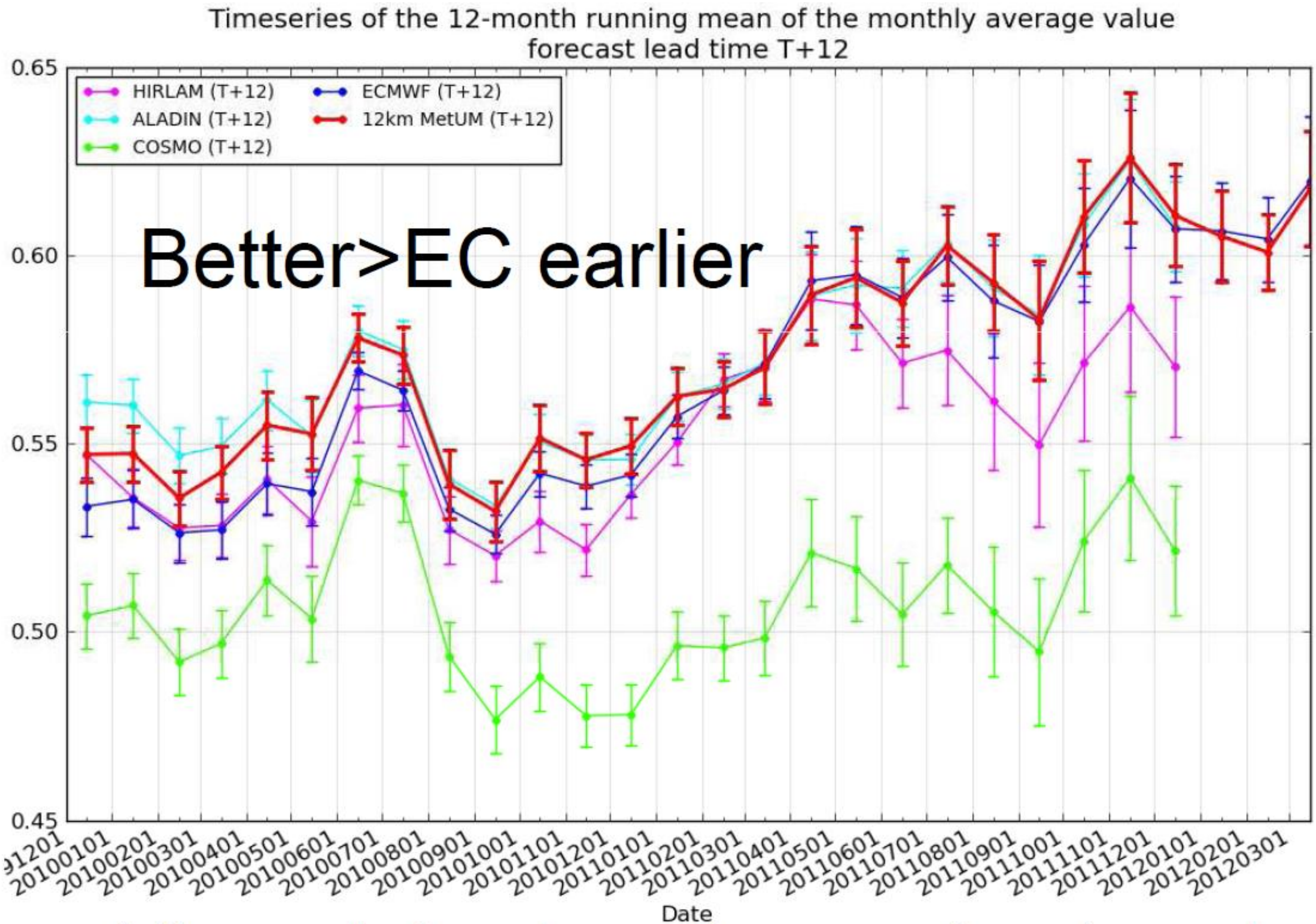
Complimentary assessment of forecast performance with climatological approaches

F.Gofa, V. Fragkouli, D.Boucouvala

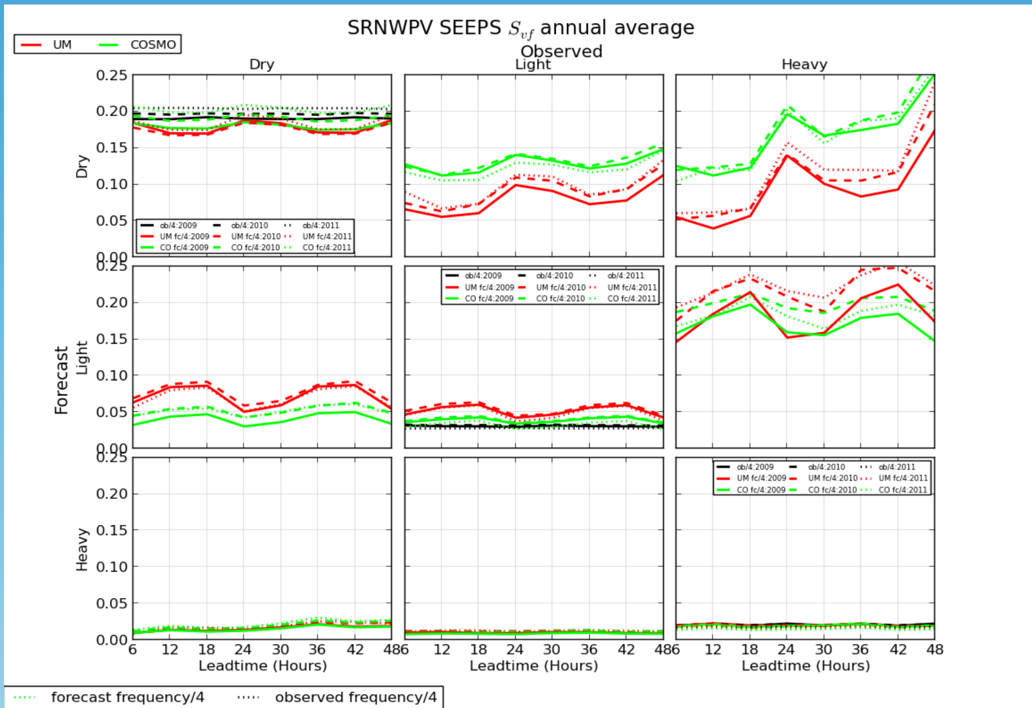
The use of SEEPS with metrics that focus on extreme events, such as the Symmetric Extremal Dependence Index (SEDI) that is adjusted to the climatological distribution of precipitation at each location, enables assessment of locally important aspects of the forecast while providing a reliable performance measure.

ΕΘΝΙΚΗ
ΜΕΤΕΩΡΟΛΟΓΙΚΗ
ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE



All models show improving trend

Considerations on SEEPS results



Note: the higher SEEPS, the worse the verification

The diagonal plots show frequency of observed and forecast in each category so no contribution in SEEPS. The off diagonal plots show how the SEEPS contributions arise.

COSMO is being penalised by:

- **Missing heavy events as u can see from both the forecast-dry (row 1 column 3) and forecast-light-precipitation (row 2, column 3) categories**
- **Missing light-precipitation events (row 2, column 3) from forecast-dry category**

THESE EVENTS SHOW UNDERESTIMATION

Also COSMO is penalised, even if to less extent, by:

- predicting light precipitation when it is observed dry (row 2, column 1)

THIS EVENT SHOWS OVERESTIMATION IN SMALL THRESHOLDS (but clearly has less weight than the others above numerically)

Positive bias of the UM (and other models) leads to a better SEEPS **as fewer heavy events are missed**. Although the UM is relatively worse at overpredicting light precipitation when dry observed (row 2, column 1), this is less of a penalty than COSMO gets for predicting dry when either light or heavy observed.

In the original paper of Rodwell on SEEPS is stated that:
"Case studies demonstrate that SEEPS is sensitive to overprediction of drizzle (our case) and failure to predict heavy large scale precipitation and incorrectly locating convective cells (again our case, where COSMO underpredicts heavy rain).

Method

Based on *North, et al. 2013* study, the objective was to apply two metrics (SEEPS and SEDI) to raingauge observations in order to evaluate if they can provide new perspectives on precipitation forecast skill at spatial scales of 7-16km.

Data set used in test

A dataset of 24h accumulated precipitation values for 18 months (**December 2013 - May 2015**) was used for 25 Greek stations. The monthly climatological values of these stations were provided by the ECMWF (process described later.)

The data was compared with precipitation forecasts from the operational regional model **COSMO 7km** and the global **IFS-ECMW**F model.

Categorical score suitable for heavy rainfall events to be used:

1. Symmetric external dependence index (SEDI)

-Equitable score

-Suitable for low base rate (rare events)

-Fixed range [-1,1], maximized as H →1 and F → 0 and minimized when H → 0 and F → 1

$$SEDI = \frac{\log F - \log H - \log(1-F) + \log(1-H)}{\log F + \log H + \log(1-F) + \log(1-H)} \quad (2)$$

where H is the hit rate,

$$H = \frac{a}{a+c} \quad (3)$$

and F is the false alarm rate,

$$F = \frac{b}{b+d} \quad (4)$$

✓***Score was added to VERSUS for easier application***

✓**Seasonal SEDI average was calculated for thresholds 10-15mm/24h as they correspond to the heavy precipitation category lower limit according to the climatology of the greek stations**

Important considerations for SEDI applications

EΘNIKH METEΩPOΛOΓIKH YΠHPEΣIA
HELLENIC NATIONAL METEOROLOGICAL SERVICE

- In order to assess the prediction of extremes separately from any model bias, forecast fields could be calibrated. However, with frequent updates of forecast system, it is not trivial to deduce the bias of subsequent model versions
- Especially true for short periods as the weather dependency of bias is usually more significant than the model version bias. No attempt for recalibration is performed in this preliminary effort, only bias can be presented simultaneously with SEDI score
- Choosing a static threshold for SEDI calculation for a large domain would require compromise between extremity and representation of stations
- In our case however it was proved that thresholds did not vary geographically so this study has to be extended to a larger domain with varying climate

MODEL	MS	1-SEEPS	t1	t2
cosmo	16682	0.804	0.2	9.267
cosmo	16754	0.446	0.2	6.933
cosmo	16614	0.931	0.2	5.1
cosmo	16641	0.659	0.2	16.767
cosmo	16743	2013 12 00015	-99.000	
cosmo	16675	0.709	0.2	5
cosmo	16648	0.684	0.2	5.1
cosmo	16650	0.841	0.2	9.933
cosmo	16734	0.822	0.2	8
cosmo	16738	0.579	0.2	8
cosmo	16667	0.968	0.2	13
cosmo	16732	0.639	0.2	6
cosmo	16749	0.567	0.2	10.933
cosmo	16684	0.469	0.2	6
cosmo	16746	0.406	0.2	9
cosmo	16622	0.967	0.2	6
cosmo	16710	0.756	0.2	8
cosmo	16643	0.481	0.2	12.133
cosmo	16627	0.309	0.2	9
cosmo	16732	0.639	0.2	8
cosmo	16732	0.639	0.2	20

Stable Equitable Error in Probability Space (SEEPS)

- Equitable
- More stable to sampling uncertainty and better for **trend detection** than other scores.
- Robust to skewed distribution because the error is measured in probability space
- Can be averaged over locations with different climates
- Adapts to assess prominent aspects of local weather.
- Inhibits hedging. It is generally not possible to reduce SEEPS without some physical insight.
- SEEPS can identify key forecasting errors including **failure to predict heavy large-scale precipitation, incorrect location of convective cells and overprediction of drizzle (decomposition).**

• **A matrix comprised of climatological probabilities for 'dry' and 'light' conditions is derived from 30 years of observations data (1980-2009) for each station.**

•The station climatology and station weights was provided by ECMWF while code to calculate SEEPS was developed at HNMS.

• The score includes three categories; 'dry', 'light precipitation' and 'heavy precipitation'. The threshold between 'dry'/'light' categories is assumed constant at 0.2mm/24h. The threshold between 'light' /'heavy' categories is calculated for every station every month.

Complimentary assessment of forecast performance with climatological approaches

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

Climatology of 24-h precipitation totals at SYNOP stations

Thomas Haiden, ECMWF

- For the computation of SEEPS that involves station climatology the weights of the SEEPS scoring matrix and the 1,2,..,99 percentiles have been computed at each station for the 30-yr period 1980-2009 (Rodwell et al. 2010).
- **The quality control procedures used in the compilation of the data are**:

  a. Rejection of individual observations by latitudinal filtering
  24-h precipitation totals have been constructed from 6-h, 12-h, and 24-h SYNOP observations disseminated via GTS. In a first step **all negative values, and all values larger than an upper limit which depends on latitude, are rejected**

  b. Rejection of stations based on percentiles
  For each month and station, **percentile values larger than 100 mm which occurred more than once were identified and, after manual inspection, discarded.**

  c. SEEPS weights are not particularly sensitive to the number of the heaviest precipitation events, and insensitive to their exact values, the resulting SEEPS score is hardly affected by such quality control.

  d. The percentiles at the high end of the distribution, on the other hand, are naturally sensitive to the heavy precipitation events, and need to be used with in mind

Stable Equitable Error in Probability Space (SEEPS)

- Equitable
- More stable to sampling uncertainty and better for **trend detection** than other scores.
- Robust to skewed distribution because the error is measured in probability space
- Can be averaged over locations with different climates
- Adapts to assess prominent aspects of local weather.
- Inhibits hedging. It is generally not possible to reduce SEEPS without some physical insight.
- SEEPS can identify key forecasting errors including **failure to predict heavy large-scale precipitation, incorrect location of convective cells and overprediction of drizzle (decomposition).**

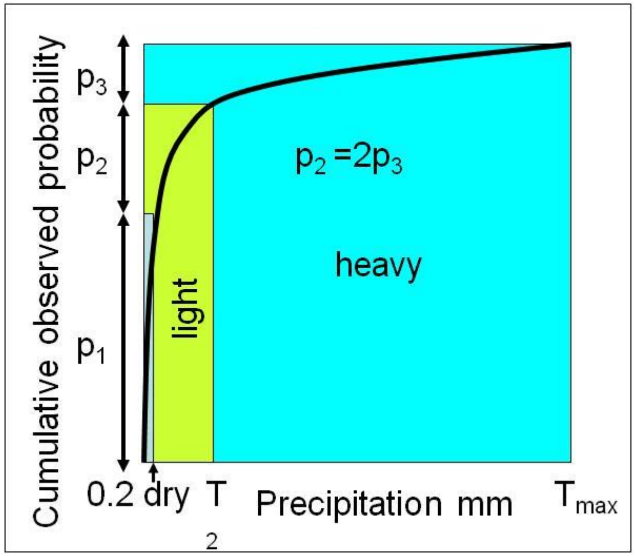
• **A matrix comprised of climatological probabilities for 'dry' and 'light' conditions is derived from 30 years of observations data (1980-2009) for each station.**

•The station climatology and station weights was provided by ECMWF while code to calculate SEEPS was developed at HNMS.

• The score includes three categories; 'dry', 'light precipitation' and 'heavy precipitation'. The threshold between 'dry'/'light' categories is assumed constant at 0.2mm/24h. The threshold between 'light' /'heavy' categories is calculated for every station every month.

Stable Equitable Error in Probability Space (SEEPS)

- Dry, light , heavy based on observed climatology (24h) at station – p_1 , p_2 , p_3
- Contingency table probabilities based on these categories
- Scoring matrix – stable, equitable
  - SEEPS=0 (perfect) , =1 (no skill - , e.g. constant)



•The SEEPS index matrix was calculated as the scalar product of the SEEPS weights matrix and the contingency table of total available model/observation pairs for each station averaged over the number of the days of the month.

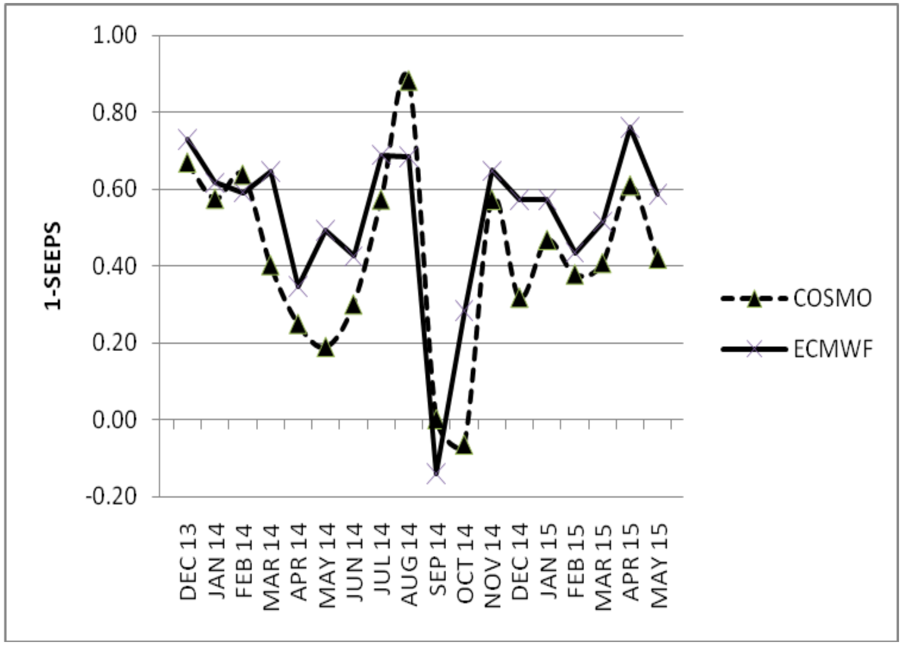
Error scoring matrix

$$\left\{ s^{S}_{vf} \right\} = \frac{1}{2} \begin{Bmatrix} 0 & \frac{1}{1-p_1} & \frac{1}{p_3} + \frac{1}{1-p_1} \\ \frac{1}{p_1} & 0 & \frac{1}{p_3} \\ \frac{1}{p_1} + \frac{1}{1-p_3} & \frac{1}{1-p_3} & 0 \end{Bmatrix}$$

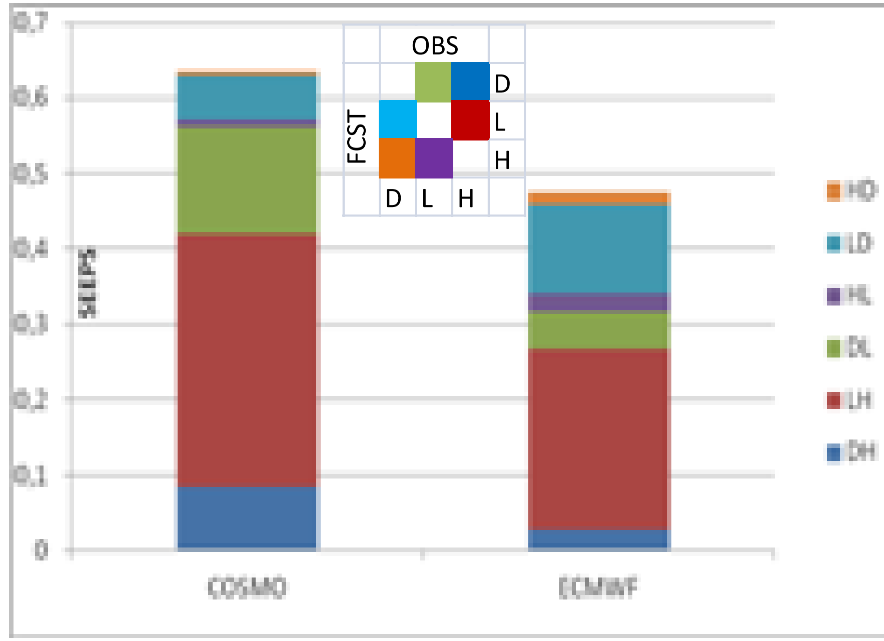
p1+p2+p3=1, p2=2p3

- Dry weather is defined as less or equal 0.2mm/24h
- The SEEPS index matrix elements are **HD** (modeled Heavy-observed Dry), **LD** (modeled Light, Observed Dry), **LH** (modeled light, observed Heavy), **DH** (modeled Dry, observed Heavy).

Monthly variation of 1-SEEPS during the observational period

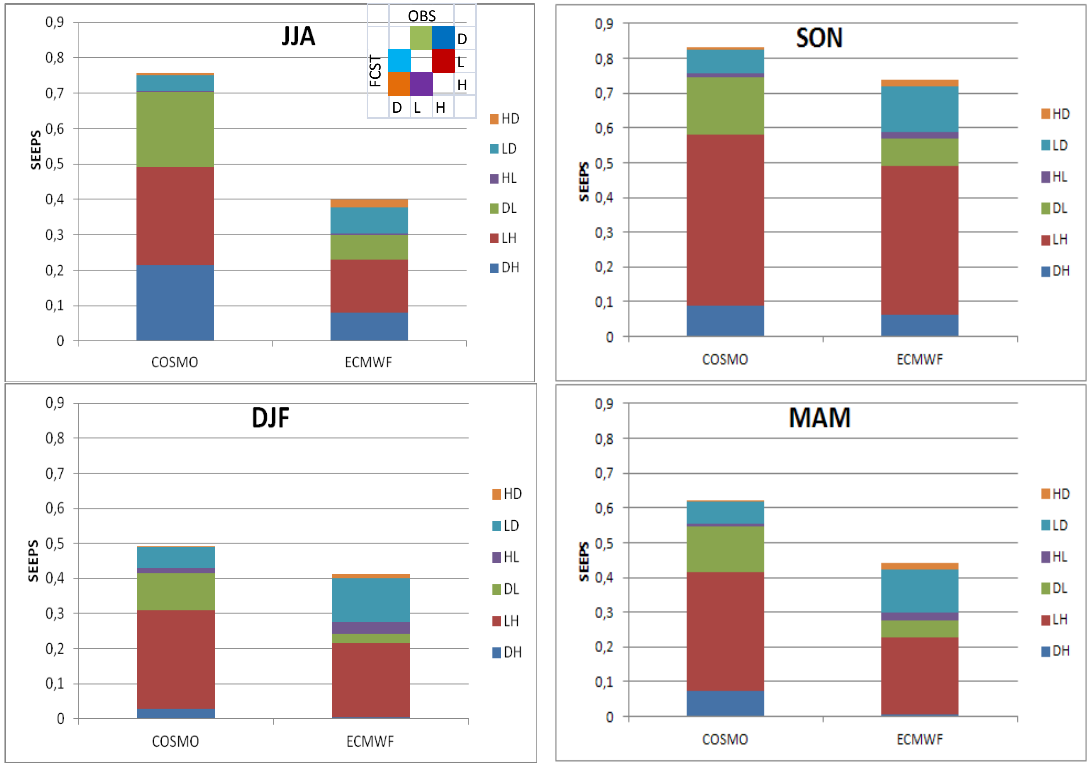


Decomposition of SEEPS for the whole period analyzed

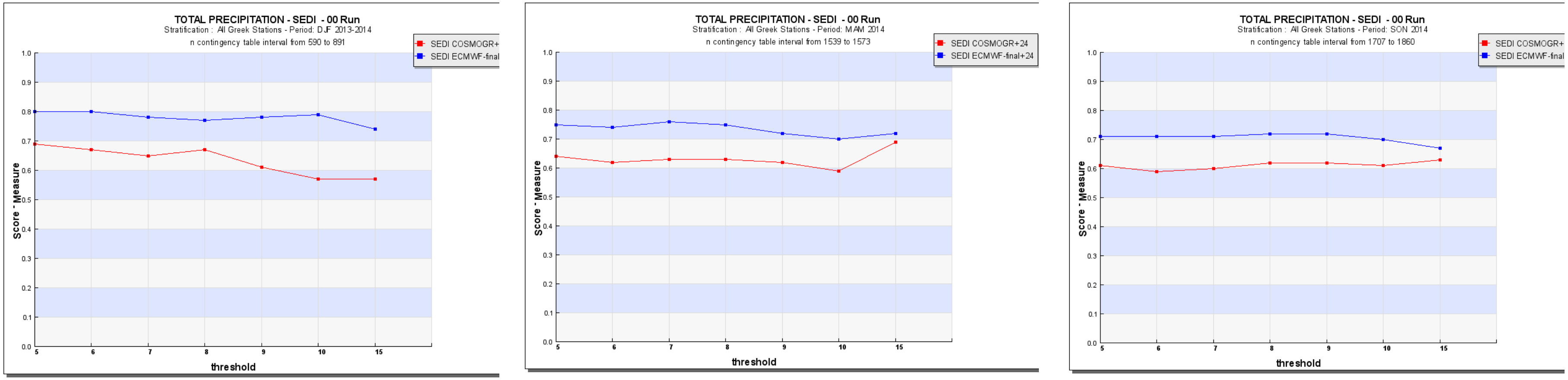


Time series for SEEPS (24h rain) exhibits poorer performance during the summer months while the ECMWF model consistently delivers better performance than the COSMO model. Both models have largest SEEPS error contribution for the ‘light’ category when ‘heavy’ was observed.

Seasonal Decomposition of SEEPS for COSMO and ECMWF models



For stations with moderate-to-dry climatologies (p1>0.5), such as Greece, predicting ‘light’ rainfall when ‘heavy’ is observed is penalized considerably more than predicting ‘light’ when ‘dry’ is observed (blue). For IFS/ECMWF, SEEPS is mainly connected with LD and LH categories, indicating that has the tendency to smooth out preci forecasts. COSMO model is penalized for LH and DL categories, leading to the conclusion that its forecast is usually ‘drier’ than that of the ECMWF model, and that the SEEPS score is strongly influenced by this attitude.

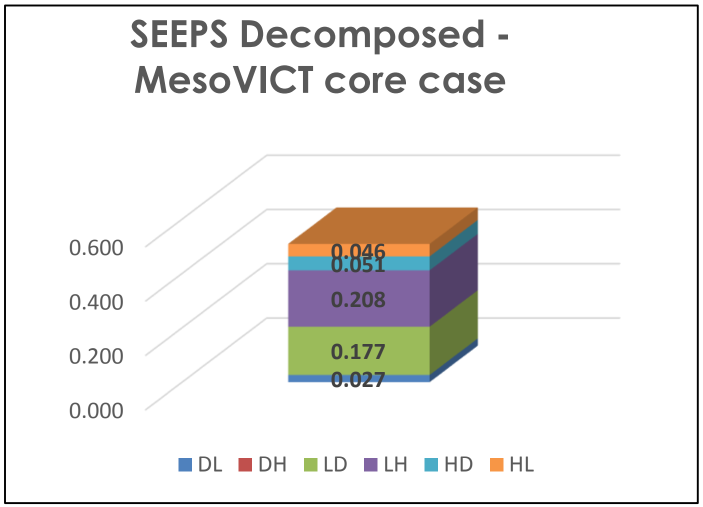


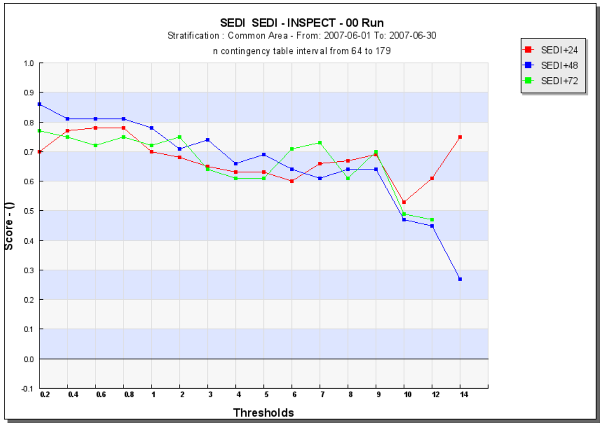
Seasonal SEDI index for DJF, MAM and SON for COSMO and ECMWF.

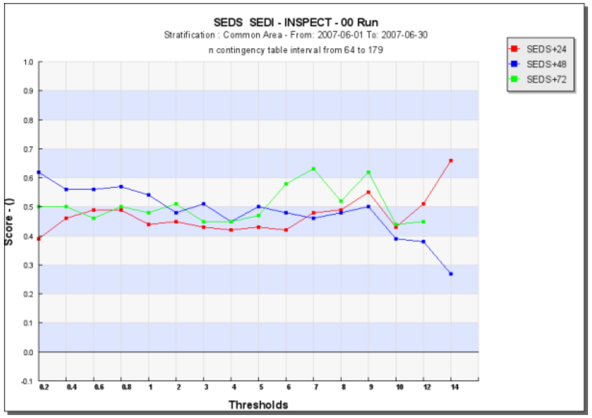
Threshold maximums are chosen according to the station climatology and here SEDI is presented only for the average of seasons and stations.
The analysis SEDI score confirmed SEEP analysis. For all seasons, the ECMWF model outperformed COSMO-GR7. SEDI values are higher in DJF and lower in SON (consistent with the SEEPS results).

Extending SEEPS application on a MesoVICT case

MODEL	MS	LAT	LON	yy	mm	pairs	WEIGHT	SEEPS	1-SEEPS	t1	t2	s11	s12	s13	s21	s22	s23	s31	s32	s33	scr11	scr12	scr13	scr21	scr22	scr23	scr31	scr32	scr33
ecmwf	11356	47.62	13.78	2007	6	3	6.637	0.898	0.102	0.2	11	0	0.898	3.594	1.128	0	2.695	1.741	0.614	0	0	0	0	0	0	2.695	0	0	0
ecmwf	11354	47.63	13.62	2007	6	3	6.7727	0	1	0.2	11	0	0.858	3.43	1.199	0	2.573	1.82	0.621	0	0	0	0	0	0	0	0	0	0
ecmwf	11357	47.73	13.45	2007	6	3	6.9939	1.11	-0.11	0.2	11	0	0.906	3.624	1.116	0	2.718	1.729	0.613	0	0	0	0	0	0	2.718	0	0.613	0
ecmwf	14648	43.35	17.8	2007	6	3	1.6142	0	1	0.2	10.067	0	1.616	6.466	0.724	0	4.849	1.281	0.557	0	0	0	0	0	0	0	0	0	0
ecmwf	16179	43.52	12.73	2007	6	3	3.5368	0	1	0.2	10	0	2.11	8.439	0.655	0	6.329	1.198	0.543	0	0	0	0	0	0	0	0	0	0
ecmwf	10963	47.48	11.07	2007	6	3	2.114	0.992	0.008	0.2	10	0	0.78	3.118	1.394	0	2.339	2.03	0.636	0	0	0	0	0	0	2.339	0	0.636	0
ecmwf	14026	46.48	15.68	2007	6	3	2.482	0.267	0.733	0.2	9.667	0	1.331	5.324	0.801	0	3.993	1.372	0.572	0	0	0	0	0.801	0	0	0	0	0
ecmwf	10946	47.72	10.33	2007	6	3	1.8218	1.145	-0.145	0.2	9.267	0	0.943	3.771	1.065	0	2.829	1.672	0.607	0	0	0	0	0	0	2.829	0	0.607	0
ecmwf	16090	45.38	10.87	2007	6	3	1.5089	0.477	0.523	0.2	9.267	0	1.659	6.634	0.716	0	4.976	1.272	0.556	0	0	0	0	1.432	0	0	0	0	0
ecmwf	16020	46.47	11.33	2007	6	3	1.6839	0.711	0.289	0.2	9.2	0	1.685	6.741	0.711	0	5.056	1.266	0.555	0	0	0	0	2.133	0	0	0	0	0
ecmwf	11150	47.8	13	2007	6	3	6.1333	1.083	-0.083	0.2	9.1	0	0.877	3.508	1.163	0	2.631	1.78	0.617	0	0	0	0	0	0	2.631	0	0.617	0
ecmwf	16172	43.47	11.85	2007	6	3	3.2807	0.214	0.786	0.2	9	0	2.277	9.106	0.641	0	6.83	1.18	0.539	0	0	0	0	0.641	0	0	0	0	0







IFS error decomposition follows the same trend as before. LD and LH are the components that contribute more greatly in the error

Next steps …

- ‘SEEPS’ has been designed to assess the general performance of precipitation forecasts including the dry/wet boundary and precipitation quantity
- SEEPS, and its decomposition into individual contributions, can be meaningfully applied to understand errors in the model’s diurnal cycle forecasts including the dry/wet boundary and precipitation quantity
- ‘SEDI’, on the other hand, has been designed to address the difficult issue of rare (extreme) events but be applied based on the climatology of each station
- Further testing is required in more climatologically diverse regions to reveal the relative advantage of SEEPS score for precipitation forecast evaluation
- SEEPS, and its decomposition into individual contributions, can be meaningfully used as trend score in Common Plots applications