Supplementary material

Details of the reservoir system model

We use weekly resolution to simulate the system and its operation for both the benchmark and the real-time optimization system (RTOS) approaches. For each reservoir (S1 and S2), the volume of stored water (*s(t+1)*) is equal to the previous week's storage (*s(t)*) plus natural and controlled inflows minus releases, evaporation and spills. The mass balance equations are:

S1: $s_{t+1} = s_t + \left(I_{S1,t} + u_{R,S1,t}\right) - \left(u_{S1,D,t} + u_{S1,R,t} + evap_t + spill_t + env_t\right)$

S2: $s_{t+1} = s_t + \left(I_{S2,t}\right) - \left(u_{S2,D,t} + evap_t + spill_t + env_t\right)$

Spills are calculated by imposing the hard constraint that the storage at next time-step should never exceed the reservoir capacity, hence they are either equal to zero or to the excess volume generated by the storage plus inflows minus outflows:

S1: $spill_t = max\left(s_t + \left(I_{S1,t} + u_{R,S1,t}\right) - \left(u_{S1,D,t} + u_{S1,R,t} + evap_t + env_t\right) - s_{max}, 0\right)$

S2: $spill_t = max\left(s_t + \left(I_{S1,t}\right) - \left(u_{S2,D,t} + evap_t + env_t\right) - s_{max}, 0\right)$

where s_{max} the reservoir storage capacity in ML. Controlled inflows and outflows (u) are limited by the real-world system capacity. Besides, pumped inflows are limited such that flow downstream of R will drop below a legal constraining value, unless using water released from S1. Evaporation fluxes (evap) are computed as the product of the reservoir surface area by the potential evaporation rate. Environmental compensation flows (env) are given by prescribed values that are kept constant over the year.

Details of the optimization

Both the release scheduling of the benchmark approach and the release and pumped inflow scheduling of the real-time optimization system (RTOS) approach are optimized using the NSGA2 genetic optimization algorithm included in the Platypus Python package (https://platypus.readthedocs.io/). The optimization decision variables are the weekly reservoir releases ($u_{S1,D}$ and $u_{S2,D}$) for both reservoir operation approaches and the weekly pumped inflows ($u_{S1,R}$) for the RTOS approach only.

For the benchmark approach the reservoir S1 operation rule curve, and not the optimizer, defines, according to the storage level and date, when pumped inflows ($u_{R,S1}$) are triggered. The optimization decision variables are the weekly reservoir releases ($u_{S1,D}$ and $u_{S2,D}$). As an optimization constraint, the storage volume for both reservoirs (S1 and S2) is set to be maximum by the end of the pumping license period window (1 April) and the objective is to minimize the sum of the pumped release ($u_{S1,D}$) energy costs:

$$\sum_{t=0}^{T} c_{R,S1}\, u_{R,S1,t} + \sum_{t=0}^{T} c_{S1,D}\, u_{S1,D,t}$$

where *c* is the pumping energy cost per ML and T the lead time in weeks.

For the RTOS approach, the optimization decision variables are the weekly reservoir releases ($u_{S1,D}$ and $u_{S2,D}$) and the weekly pumped inflows ($u_{S1,R}$) and the optimization objective is to minimize the following objective functions:

1) Average of the difference between the storage capacity and storage volume by 1 April for the two reservoirs (S1 and S2):

$$\frac{(s_{S1,max} - s_{S1,T}) + (s_{S2,max} - s_{S2,T})}{2}$$

where *s* is the reservoir storage volume in ML and s_{max} the reservoir storage capacity in ML.

2) Sum of the pumping energy costs (only applied on the multi-objective optimization of the RTOS approach):

$$\sum_{t=0}^{T} c_{R,S1}\, u_{R,S1,t} + \sum_{t=0}^{T} c_{S1,D}\, u_{S1,D,t}$$

where *c* is the pumping energy cost per ML and T the lead time in weeks.

For both operation approaches the optimization consisted of 100,000 runs per iteration and the population size for the multi-objective optimization of the RTOS approach was 20.

Supplementary figures

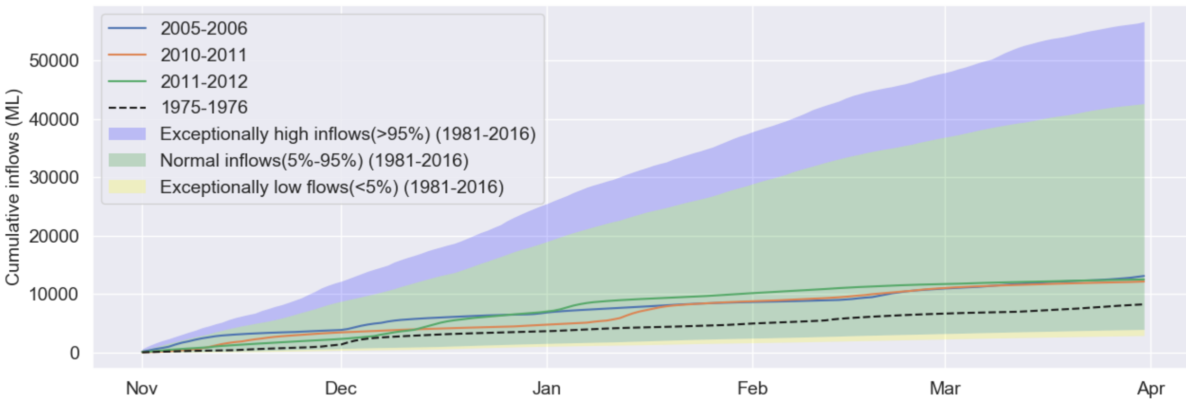


Figure 8 Cumulative inflows to the S1 reservoir in the worst-case scenario (1975-1976) and in the three driest years (2005-2006, 2010-2011 and 2011-2012) of the period used for the simulation of the RTOS (2005-2016). Only data relative to the pumping licence window (Nov to Apr) are shown. Shaded areas show the weekly inflow distribution calculated on the period used for the forecast bias correction and ESP generation (1981-2016). Notice that the three driest years are relatively close to the worst-case scenario.

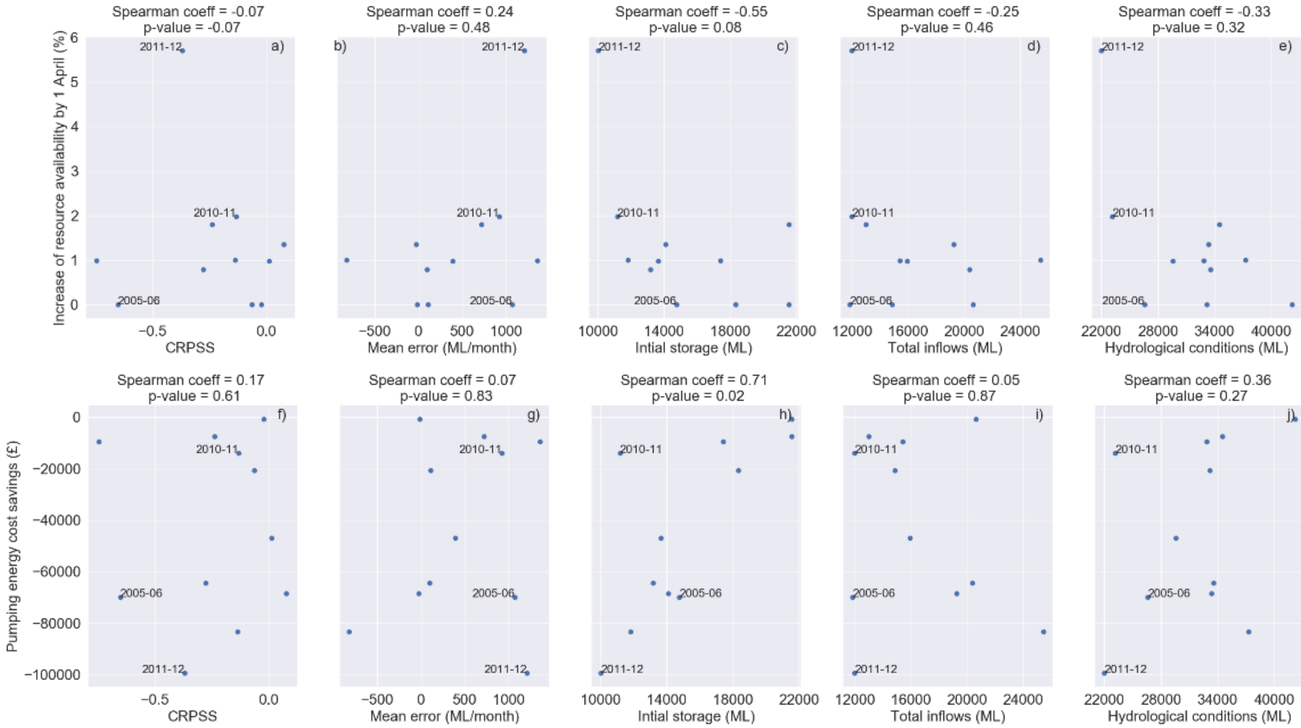
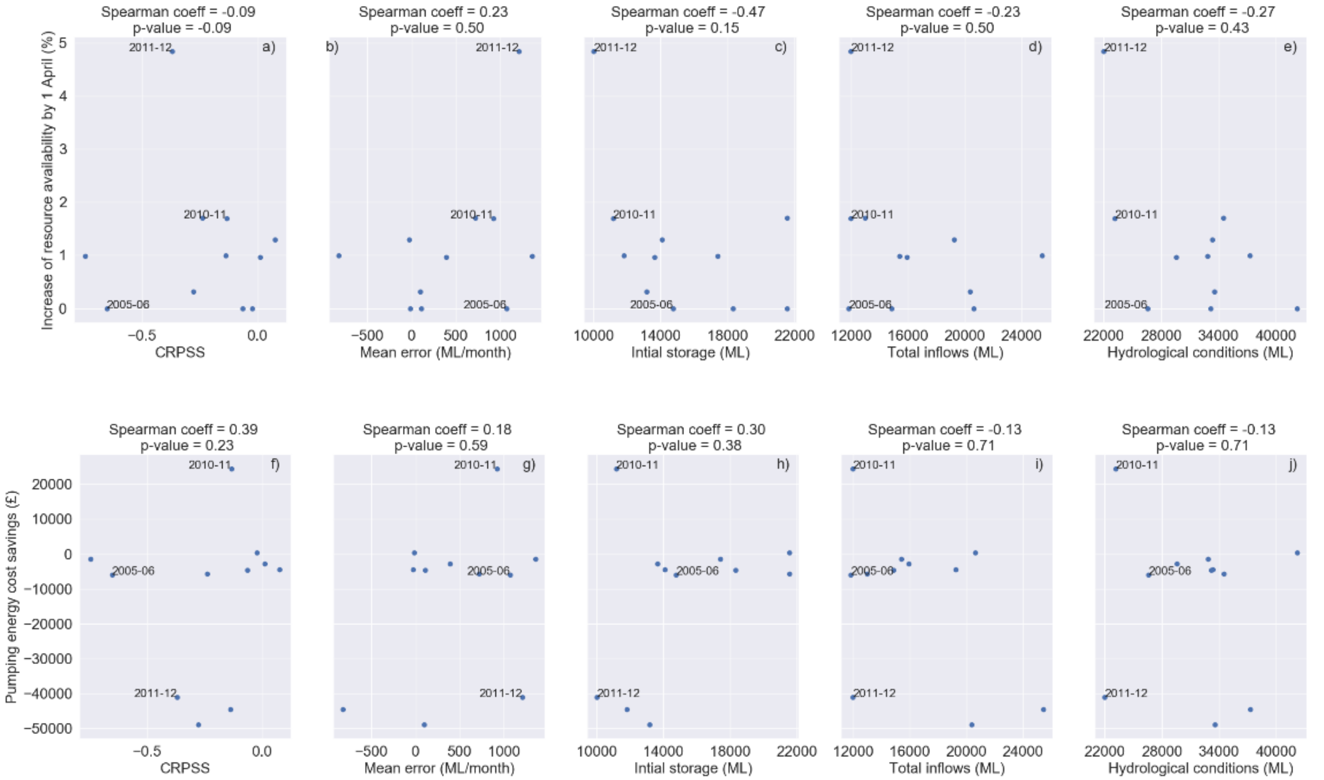


Figure 9 Ensemble streamflow prediction (ESP) for the "resource availability only" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.



745 **Figure 10 Ensemble streamflow prediction (ESP) for the "resource availability prioritised" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.**

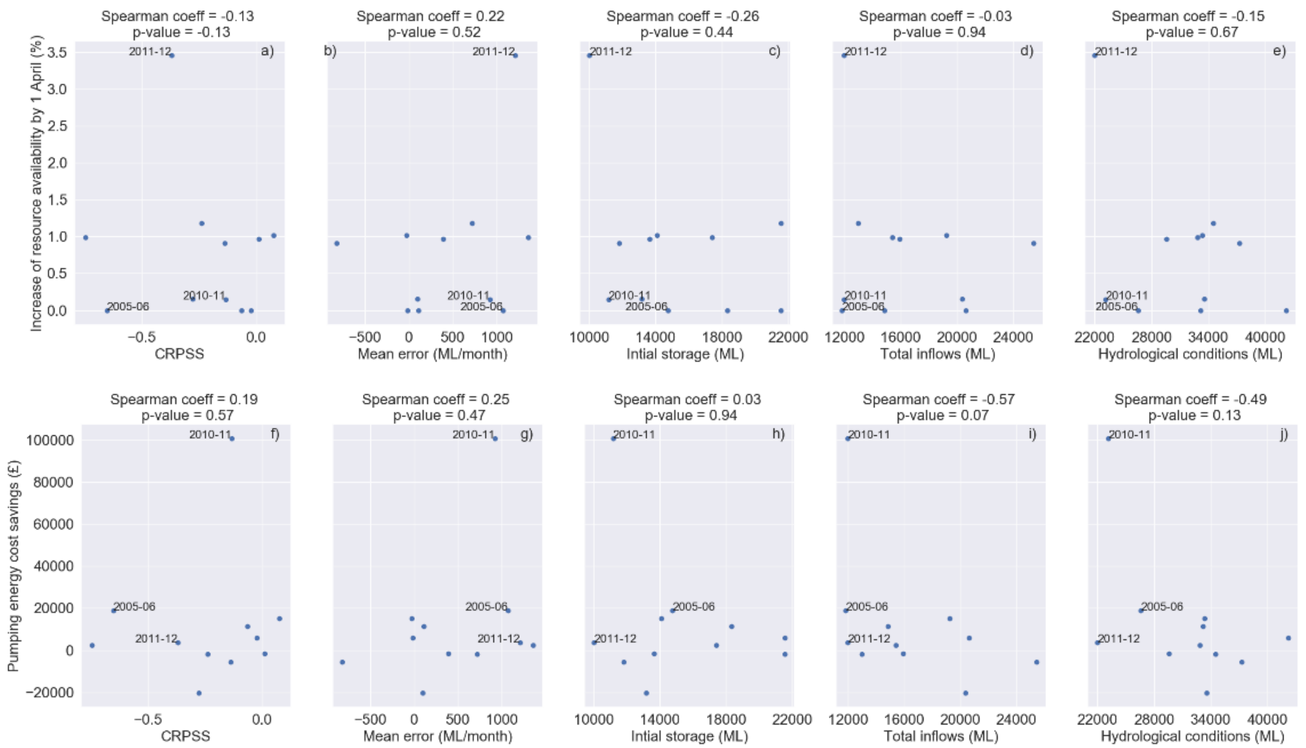


Figure 11 Ensemble streamflow prediction (ESP) for the "balanced" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.

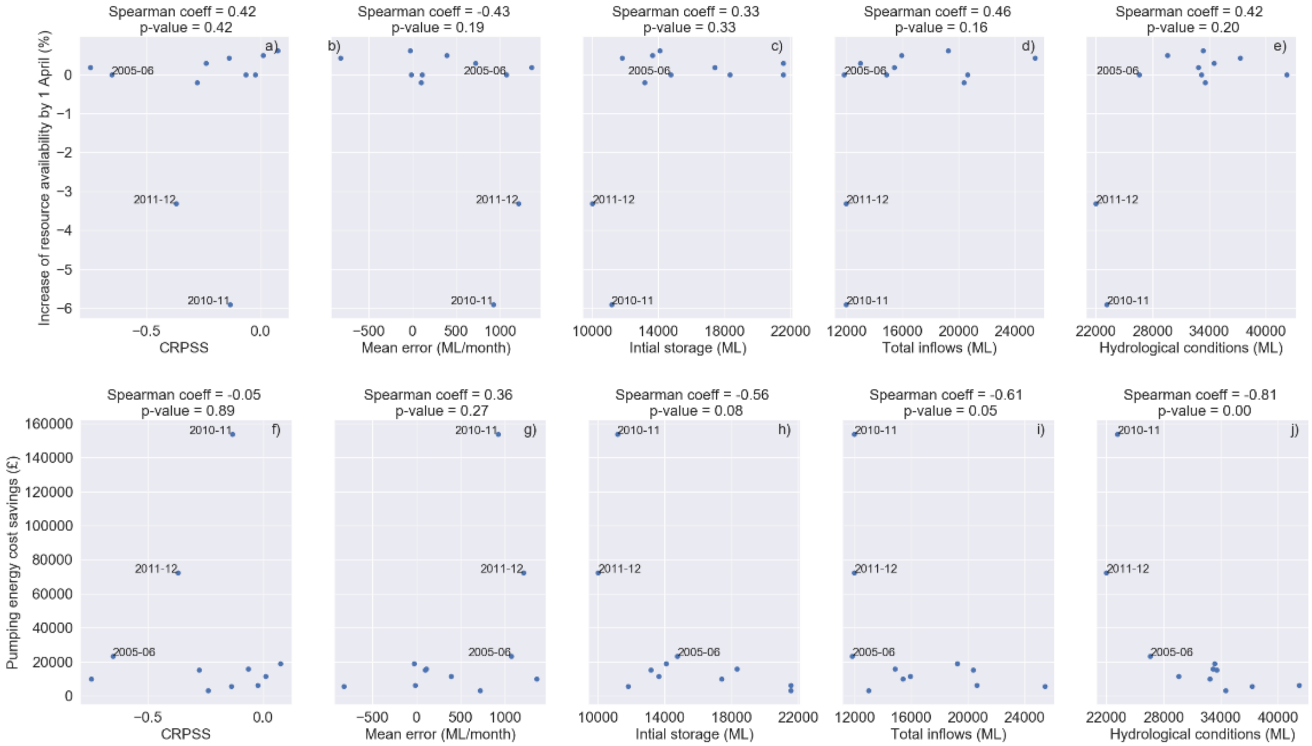


Figure 12 Ensemble streamflow prediction (ESP) for the "pumping savings prioritised" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.

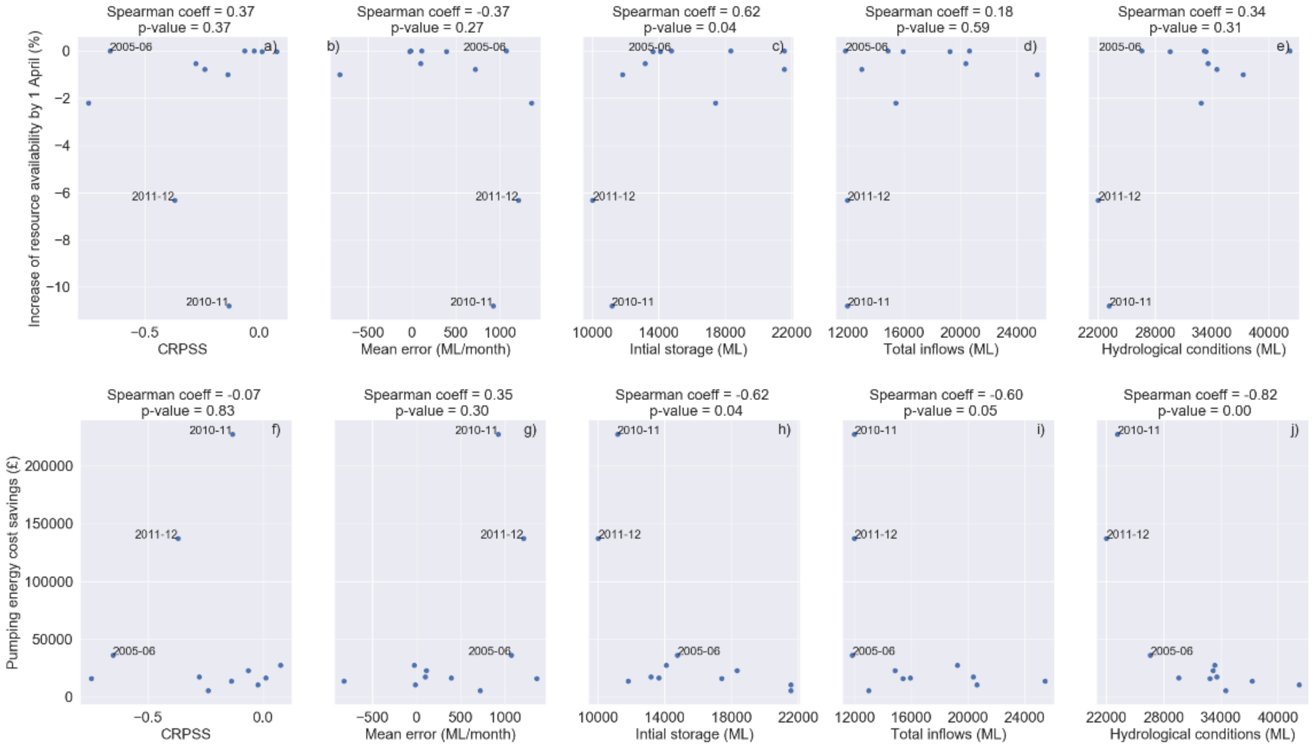


Figure 13 Ensemble streamflow prediction (ESP) for the "pumping savings only" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.

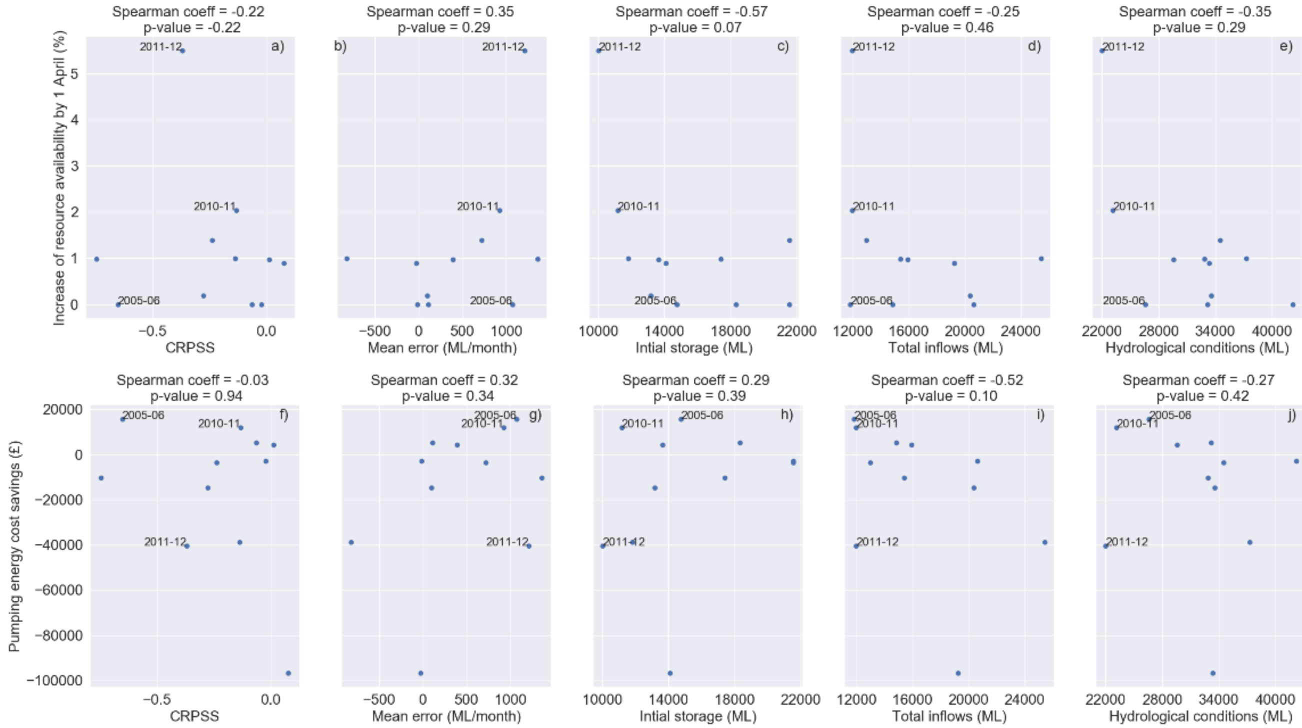


Figure 14 Bias corrected forecast ensemble (DSP-corr) for the "resource availability only" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.

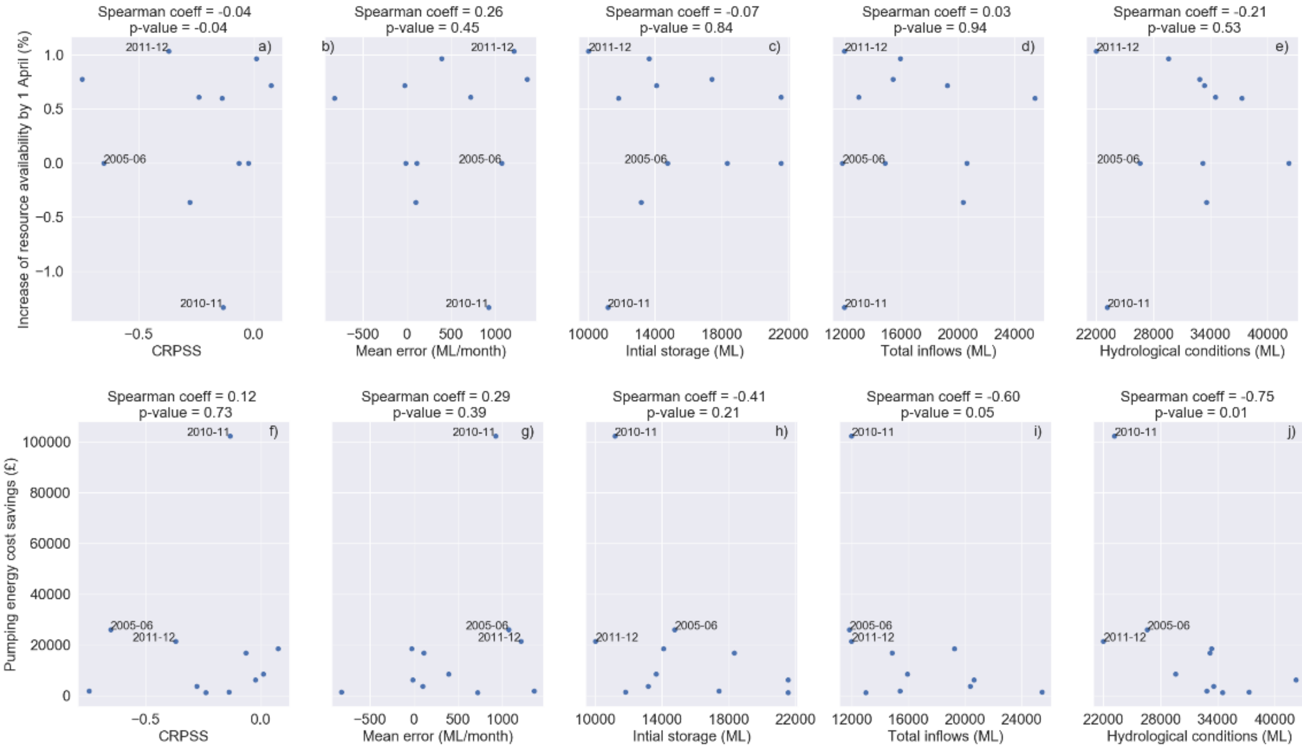


Figure 15 Bias corrected forecast ensemble (DSP-corr) for the "balanced" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.

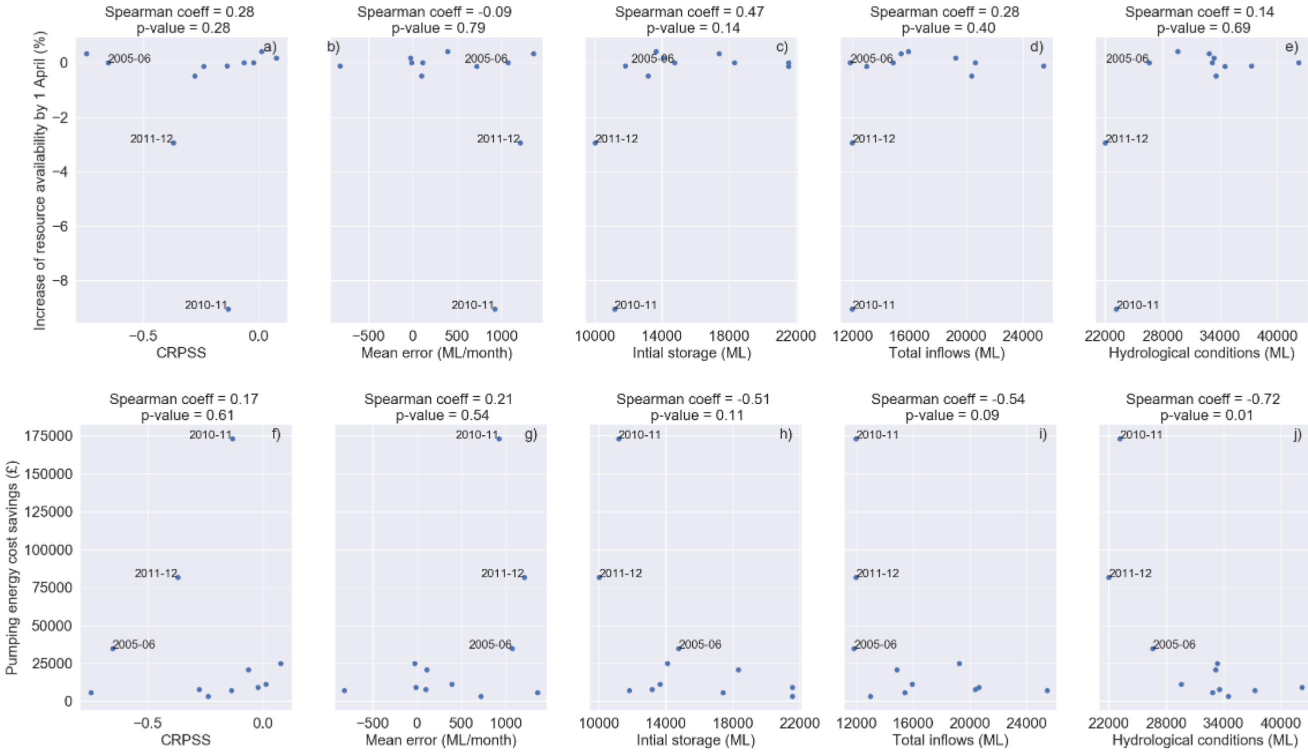


Figure 16 Bias corrected forecast ensemble (DSP-corr) for the "pumping savings prioritised" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.

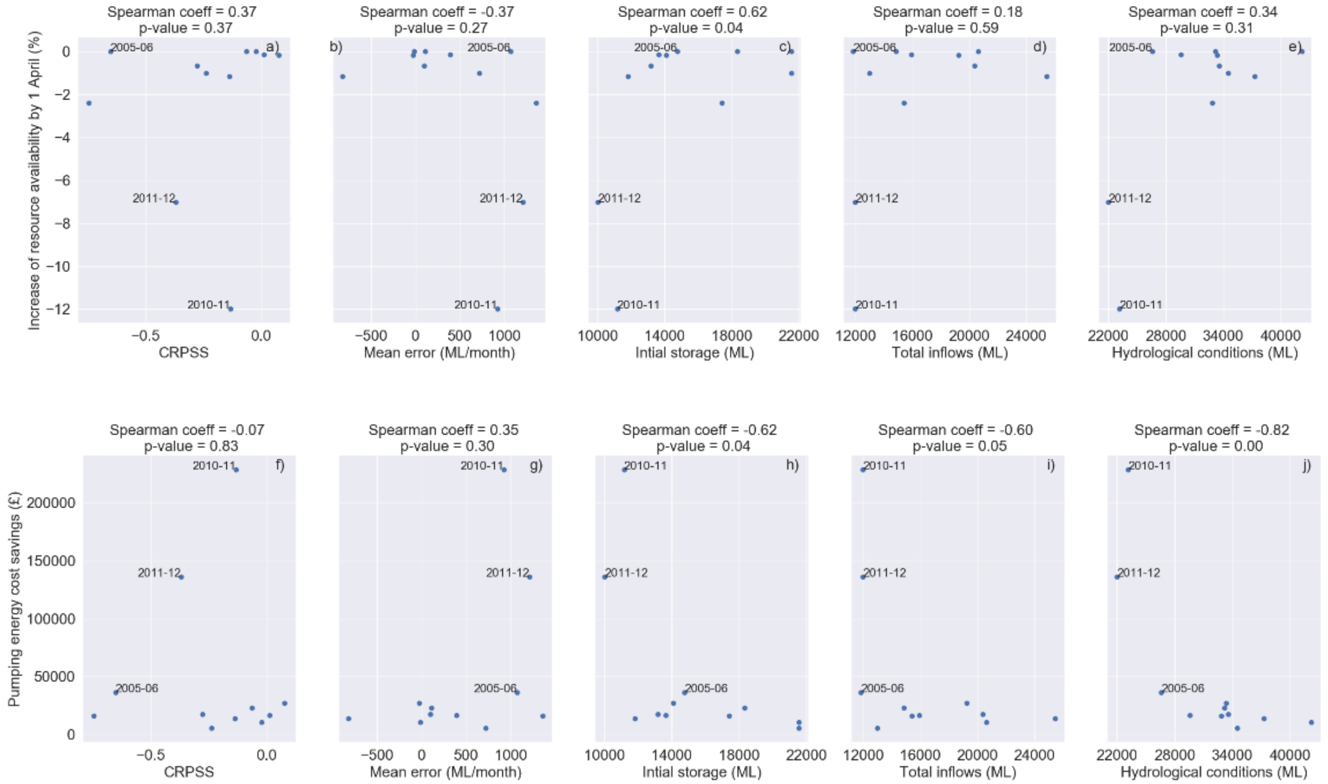


Figure 17 Bias corrected forecast ensemble (DSP-corr) for the "pumping savings only" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.