Using machine learning to enhance skill of subseasonal-to-seasonal (S2S) temperature forecasts

Aheli Das[1], David Brayshaw[1], John Methven[1], Thomas Frame[1], Christopher O' Reilly[1], Shivkumar Sharma[2], Jake Mammatt[2], Shane Fox[2]

[1]Department of Meteorology, University of Reading, UK

[2]British Gas, Centrica, UK

University of Reading

British Gas

EGU26 GENERAL ASSEMBLY

X5.22
EGU26-11985

Online machine learning: Sequential learning algorithm (SLA)

- SLA is a relatively new algorithm which has been used to obtain skillful S2S predictions at time-scales of 2 weeks to a season [1].
- SLA optimally combines different 'experts' or predictors using dynamic weights while minimizing quantile loss [2].
- The weights are updated every time step to favour the good 'expert's, i.e., 'expert's with low losses will get higher weights in the next step and vice-versa.
- At each time-step, forecast *y* at time *t* equals the weighted sum of experts *e*: $y_t = \sum w_t e_t$

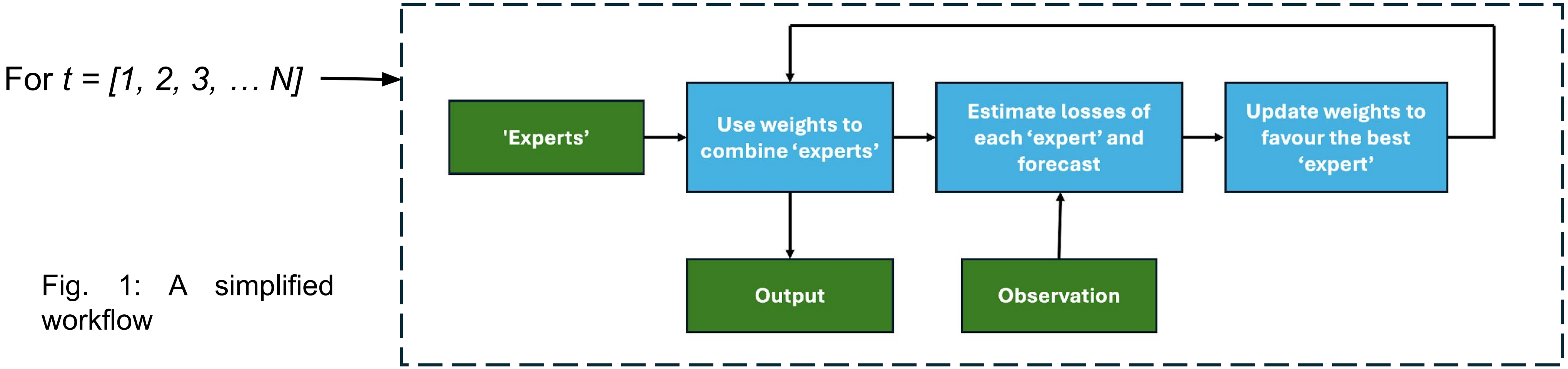

Fig. 1: A simplified workflow

[1] Gonzalez et al. (2021)
[2] Berrisch & Ziel (2021)

Data and methods

Predictors:

Forecast experts:

ECMWF S2S T2m quantiles from 2003-2022

Climatological experts:

ERA5 T2m climatological quantiles adjusted for warming from 1970 to year prior to forecast initialisation

Target:

Weekly mean ERA5 T2m at 1-4 weeks lead time

Time period:

Training/spin-up: 2003-2007

Testing: 2008-2022

Verification metrics:

Error/Bias

Continuous ranked probability skill score (CRPSS)

- Probabilistic weekly mean T2m forecasts are generated using SLA at 1-4 weeks lead time.
- Raw ECMWF S2S ensembles are corrected via. linear scaling using leave-one-out-cross-validation prior to calculating experts.

$$y'_{i,m} = y_{i,m} \frac{\bar{o}}{\bar{y}}$$

$y'_{i,m}$ = calibrated m^{th} ensemble for i^{th} time step

$y_{i,m}$ = raw m^{th} ensemble for i^{th} time step

$\bar{o}$ = mean climatology

$\bar{y}$ = ensemble mean of raw forecasts

- Total 22 experts used as SLA inputs:

  11 Forecast experts:

  [*for_min, q10,q20...,q90, for_max*]

  11 Clim experts:

  [*clim_min, q10,q20...,q90, clim_max*]

SLA working principle

For each expert *e* at time *t:*

Estimate instantaneous regret r_t i.e. difference between forecast error and expert error

Estimate range and variance of regret E_t and V_t

Estimate learning rate $\eta_t = f(w_0, E_t, V_t)$

Estimate cumulative regret $R_t = f(r_t, \eta_t)$

Update weights $w_t = f(\eta_t, R_t)$

- Probabilistic predictions are generated by iterating the equations over a defined probability grid τ

$$\tau = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$$

- l is a quantile loss calculated as

$$l = \begin{cases} \tau.\,|x - y| \; if \; y < x \\ (1 - \tau).\,|x - y| \; if \; y > x \end{cases}$$

y is the forecast and *x* is the observation.

The objective is to leverage the capabilities of SLA to generate skilful T2m predictions at weeks 1-4.

Quantile loss

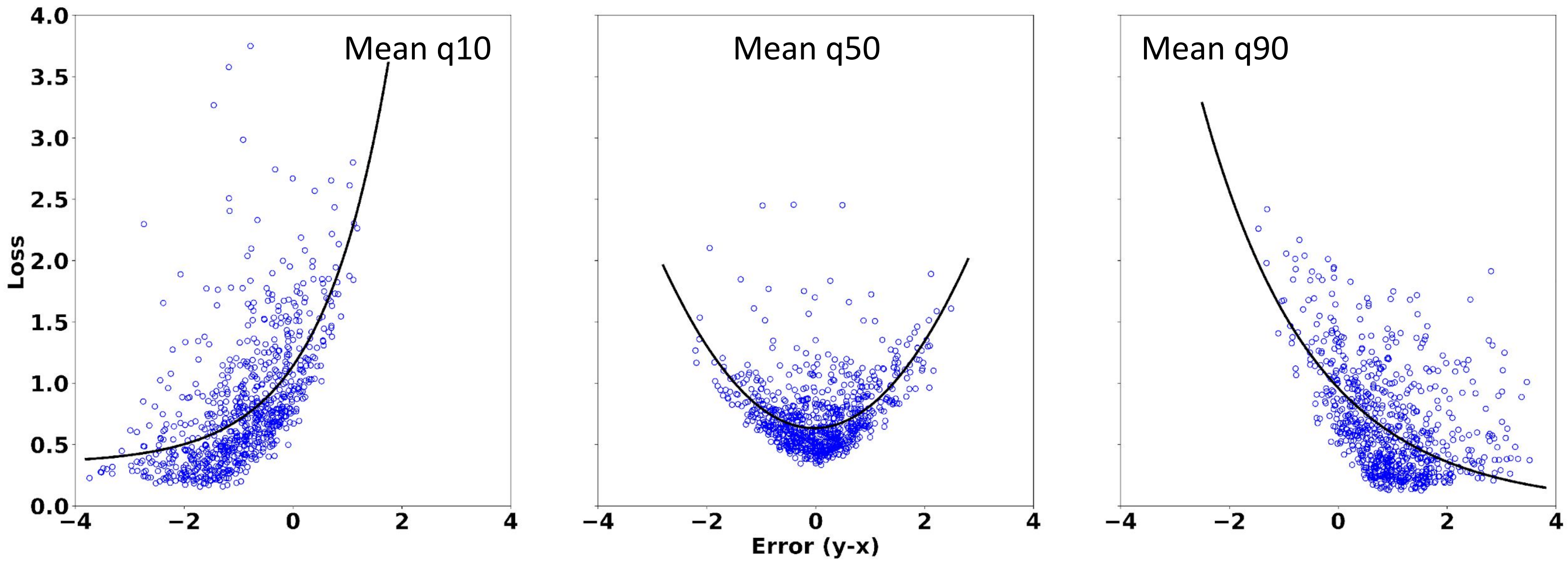

Fig. 2: SLA error vs. mean expert loss for 10th, 50th and 90th percentiles respectively for week 1 during the test period 2008-2022. Black line is the line of best fit.

- Quantile loss curves are an important diagnostic tool for probabilistic forecasting.
- For 10th (90th) percentile, overpredictions (underpredictions) are heavily penalized; resulting in low (high) predictions. Symmetric only for median or 50th percentile.
- Strongly suggests SLA's capability to produce probabilistic predictions of mean T2m.

Expert weight evolutions

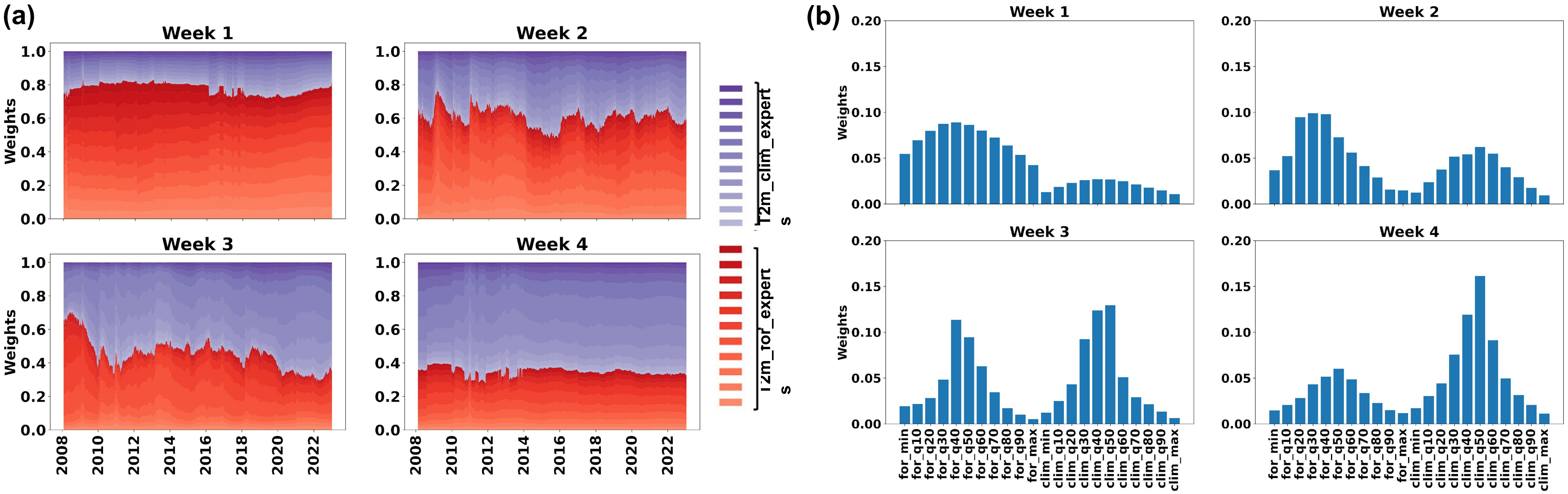

Fig. 3(a): Weight evolutions of 'experts' for the median or q50 T2m prediction during the test period 2008-2022 for weeks 1-4. Lighter to darker shadings indicate lower to higher quantiles. (b) Mean weights of each expert quantile during the similar period for weeks 1-4.

- SLA benefits from initial conditions in T2m forecasts in weeks 1-2. So, has large share of weights on forecast experts.
- Conversely, SLA has large share of weights on climatological experts in weeks 3-4, as initial conditions in forecasts can no longer provide useful information.
- Seasonality in weights is driven by the cumulative past performances of the individual experts.

Forecast verification

University of Reading

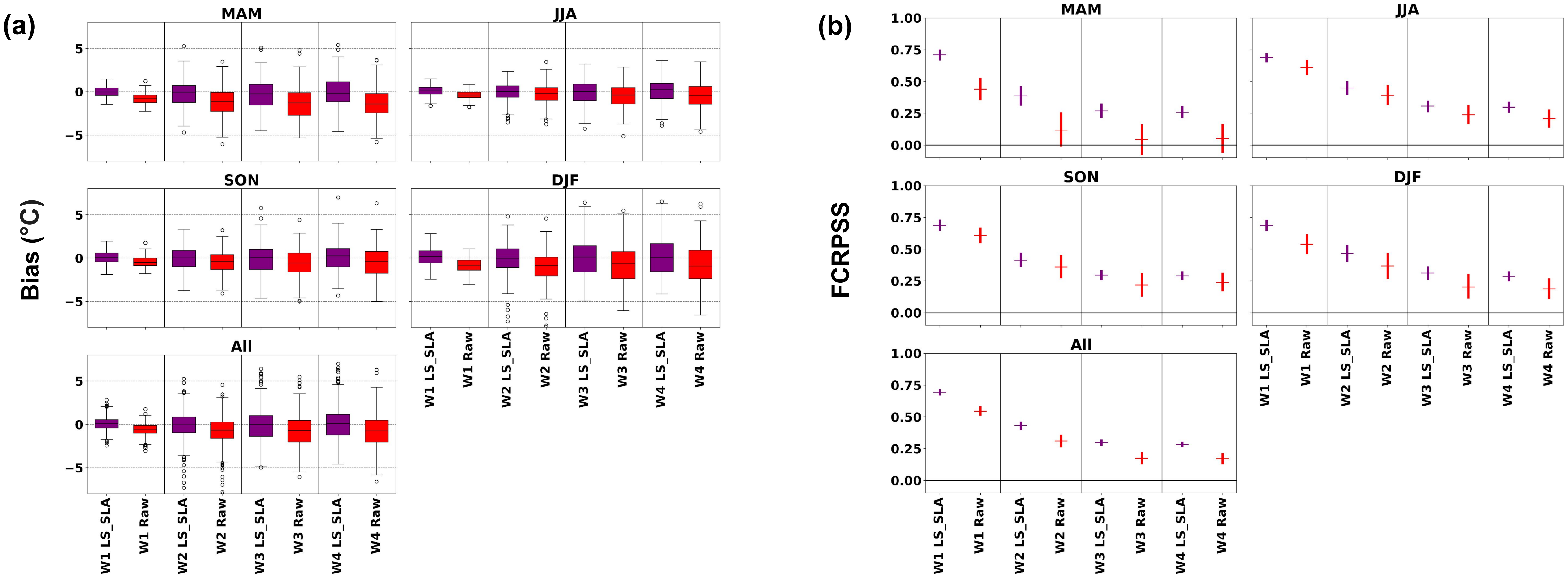

Fig. 4: (a) Boxplots of bias of q50 T2m predictions from SLA and raw forecasts. (b) FCRPSS of all Q10-Q90 SLA T2m predictions and raw ensemble forecasts at weeks 1-4 in the test period.

- Raw T2m forecasts overall have a negative bias at all weeks; SLA improves this and the median bias is now centered on zero.
- Improvements range from 27-39% in weeks 1-2 and 66-70% in weeks 3-4.

Summary and future work

University of Reading

- Overall, SLA is able produce skilful deterministic and probabilistic T2m forecasts at S2S time-scales over raw uncalibrated forecasts.
- Unlike traditional bias-correction method, SLA can exploit information from climatology after week 2 when initial condition information from raw forecasts deplete.
- Weight evolutions shows how SLA learns to favour different experts across lead times, providing interpretability of its working mechanism.

- SLA weights can also be used to obtain insights on predictability.
- The ‘online’ nature makes SLA suitable for operational uses as it can seamlessly transition between different versions of the NWP model.
- Can be expanded to blend in multiple NWP model outputs or multiple meteorological variables to produce multimodel forecasts or forecast any composite weather variable.

Thank you

Contact

aheli.das@reading.ac.uk