Data-driven derivation of equations for the evolution of transport in turbulent flow

C.J. Wareing, A. Roy, S.M. Tobias

School of Mathematics, University of Leeds, Leeds LS2 9JT, UK. *C.J.Wareing@leeds.ac.uk*

Summary

We present results of DNS of turbulent fluid dynamics coupled with machine learning techniques to derive new equations for the evolution of transport in turbulent flows. We examine Rayleigh-Bénard convective turbulence with the aim to learn the statistics of unresolved scales for turbulent parameterization. Following the approach of Garaud et al. 2010 [1] in order to perform a comparison to their result, we seek a closure model for the transport of entropy and momentum intended for application to rotating stellar convective regions. We use the Dedalus framework [2,3] for spectrally solving differential equations to generate an extended time-series of two-dimensional DNS data at Rayleigh numbers of 10^6, 10^8 and 10^{10}. We then use the data-driven Sparse Identification of Nonlinear Dynamics (SINDy) machine-learning algorithm [4] to discover the form of the governing equations paying particular attention to capturing any difference between the bulk and boundary layers. In future work, we will consider different approaches to the machine-learning question (e.g. SPIDER and Bayesian methods) and seek to confirm the turbulence closure coefficients of Garaud et al. 2010 as well as consider differences in the coefficients relating to different regions of the flow.

Physical Model & Governing Equations

- 2D horizontally-periodic Rayleigh-Bénard convection.
- Non-dimensionalised using the box height and freefall time.
- Rayleigh numbers, $R = 10^6$, 10^8 & 10^{10}. Prandtl number, $P = 1$.
- Stress-free boundary condition at top $z = 0$ and bottom $z = 1$.
- Boussinesq approximation.
- Temperature $\Theta(z=0)=1$, $\Theta(z=1)=0$.

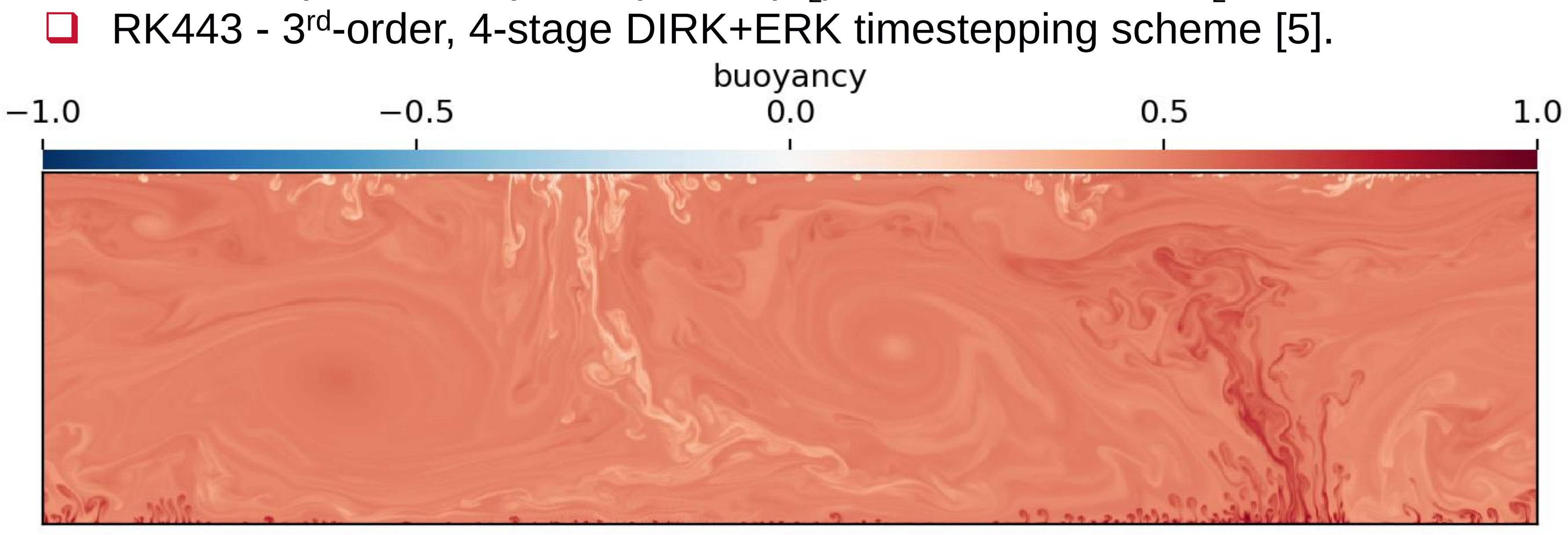
$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\nabla\sigma + \Theta\hat{z} + \sqrt{\frac{P}{R}}\nabla^2\boldsymbol{u}, \quad (1)$$

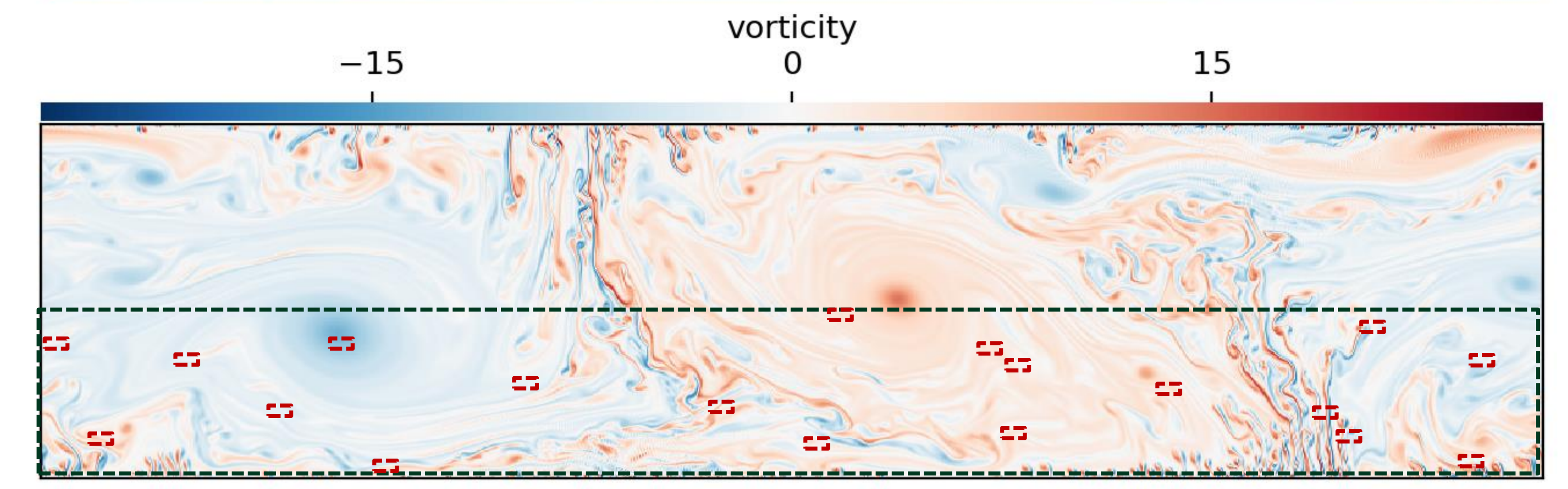
$$\frac{\partial \Theta}{\partial t} + (\boldsymbol{u}\cdot\nabla)\Theta = \frac{1}{\sqrt{PR}}\nabla^2\Theta, \quad (2)$$

$$\nabla\cdot\boldsymbol{u} = 0, \quad (3)$$

Numerical technique

- Dedalus v3 flexible framework for solving PDEs using spectral methods [2].
- Open-source, widely-used, well supported [3].
- 1024 real space Fourier points (N_x) in the x direction, $L_x = 4$.
- 384 real space Chebyshev points (N_z) in the z direction, $L_z = 1$.
- RK443 - 3rd-order, 4-stage DIRK+ERK timestepping scheme [5].





- Snapshot at t=48.5 of resolved 2D convection, Ra = 10^{10}.

Theoretical model

- Garaud et al. [1] retained exact forms of the left-hand sides, developing equations for the triple correlation terms and proposed simple closures for the right-hand sides by splitting mean and fluctuating parts $u_i = \bar{u}_i + u_i'$ with $\bar{u_i'} = 0$, of the form:

$$(\partial_t + \bar{u}_k\partial_k)\bar{R}_{ij} + \bar{R}_{ik}\partial_k\bar{u}_j + \bar{R}_{jk}\partial_k\bar{u}_i + \alpha(\bar{F}_i g_j + \bar{F}_j g_i) - \nu\partial_{kk}\bar{R}_{ij} = -\frac{C_1}{L}\bar{R}^{1/2}\bar{R}_{ij} - \frac{C_2}{L}\bar{R}^{1/2}\left(\bar{R}_{ij} - \frac{1}{3}\bar{R}\delta_{ij}\right) - \nu\frac{C_\nu}{L^2}\bar{R}_{ij}, \quad (4)$$

$$(\partial_t + \bar{u}_j\partial_j)\bar{F}_i + \bar{R}_{ij}\partial_j\bar{\Theta} + \bar{F}_j\partial_j\bar{u}_i + \alpha\bar{Q}g_i - \tfrac{1}{2}(\nu+\kappa)\partial_{jj}\bar{F}_i = -\frac{C_6}{L}\bar{R}^{1/2}\bar{F}_i - \frac{1}{2}(\nu+\kappa)\frac{C_{\nu\kappa}}{L^2}\bar{F}_i, \quad (5)$$

$$(\partial_t + \bar{u}_i\partial_i)\bar{Q} + 2\bar{F}_i\partial_i\bar{\Theta} - \kappa\partial_{ii}\bar{Q} = -\frac{C_7}{L}\bar{R}^{1/2}\bar{Q} - \kappa\frac{C_\kappa}{L^2}\bar{Q}, \quad (6)$$

where $\bar{R}_{ij} = \overline{u_i' u_j'}$, $\bar{F}_i = \overline{\Theta' u_i'}$ and $\bar{Q} = \overline{\Theta'^2}$.

- Assuming variation only with z, they obtain a set of coefficients for 3D Rayleigh-Bénard convection as follows:

$$C_1 \simeq 0.4, \quad C_2 \simeq 0.6, \quad C_\kappa = 2 \pm 0.2.$$
$$C_\nu = 12 \pm 1, \quad C_6 = 1.4 \pm 0.1, \quad (7)$$
$$C_{\nu\kappa} = 6 \pm 0.5, \quad C_7 = 1.4 \pm 0.1.$$

SINDy

- Discovering governing equations from data by **S**parse **I**dentification of **N**onlinear **Dy**namical (SINDy) systems [4]
- Figure 1 from Brunton et al. [4] gives a demonstrative overview:-

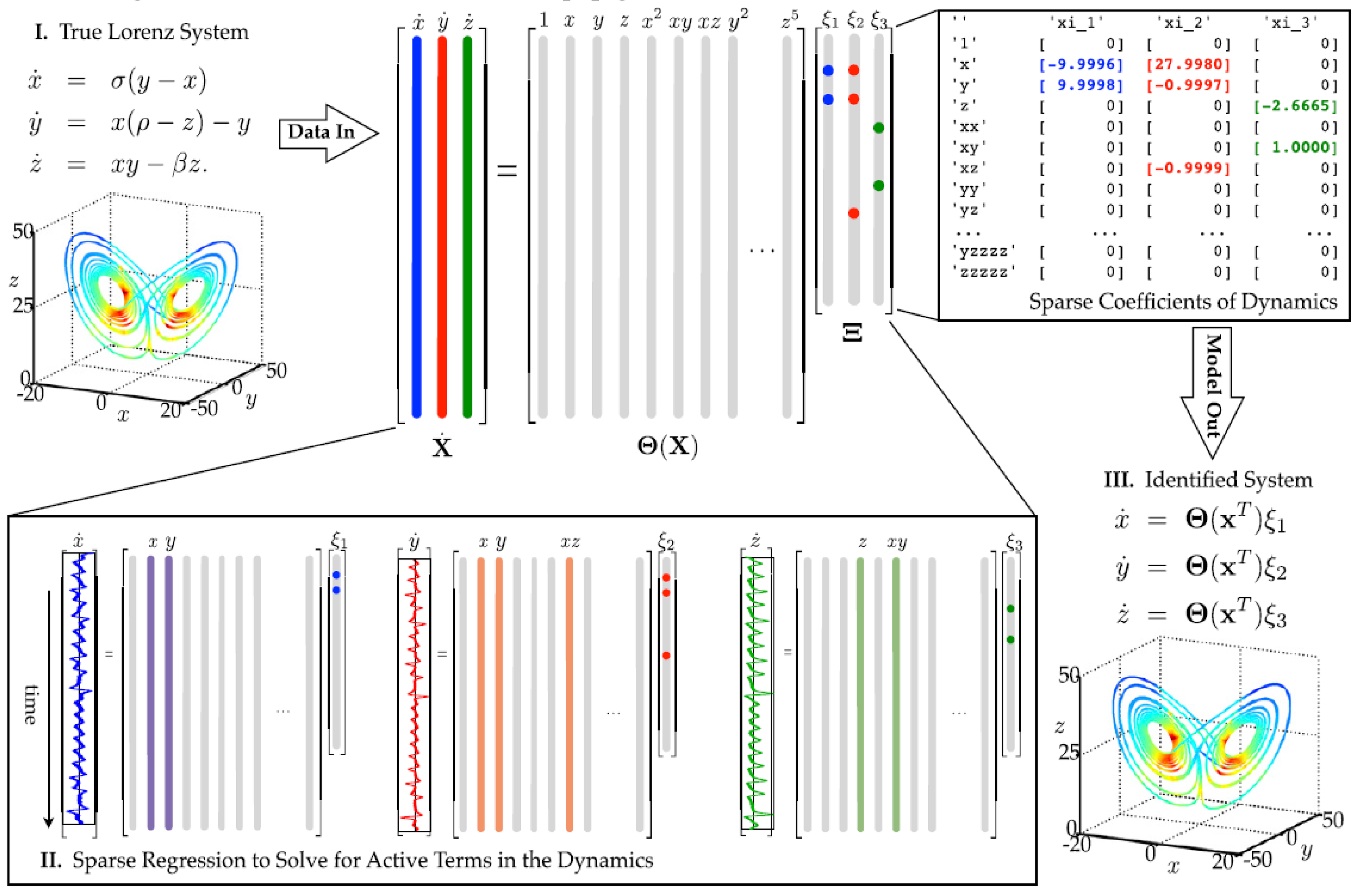


Fig. 1. Schematic of the SINDy algorithm, demonstrated on the Lorenz equations. Data are collected from the system, including a time history of the states X and derivatives Ẋ; the assumption of having Ẋ is relaxed later. Next, a library of nonlinear functions of the states, Θ(X), is constructed. This nonlinear feature library is used to find the fewest terms needed to satisfy Ẋ = Θ(X)Ξ. The few entries in the vectors of Ξ, solved for by sparse regression, denote the relevant terms in the right-hand side of the dynamics. Parameter values are $\sigma = 10$, $\beta = 8/3$, $\rho = 28$, $(x_0, y_0, z_0)^T = (-8,7,27)^T$. The trajectory on the Lorenz attractor is colored by the adaptive time step required, with red indicating a smaller time step.

Results: recovering the governing equations

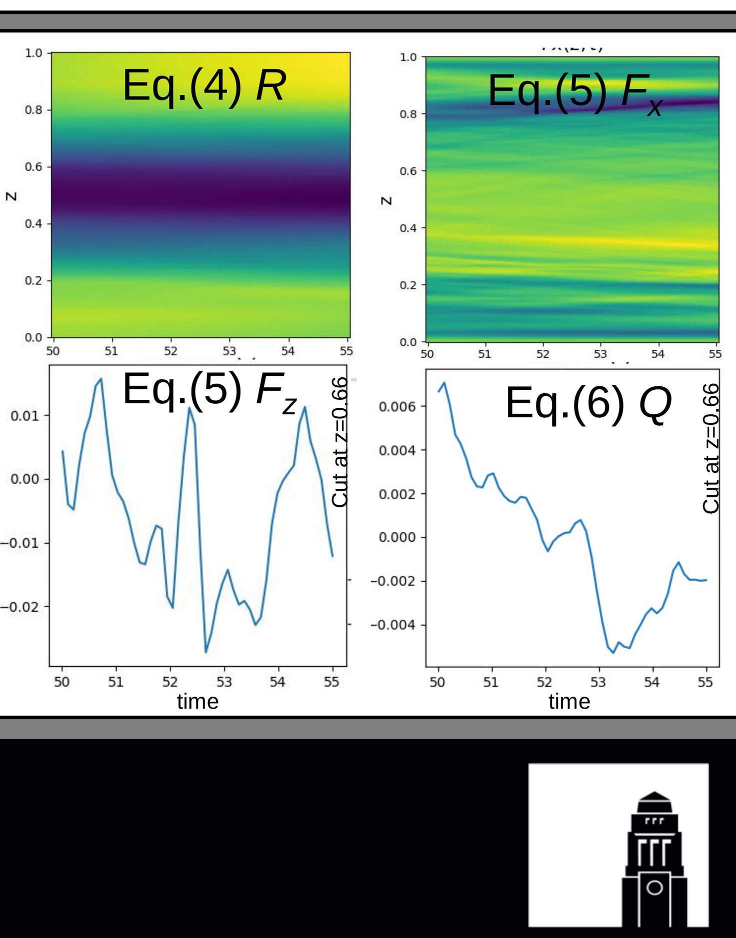
- Applying SINDy with constraints [(1) restrictions on the number of diffusive terms, (2) the incompressibility condition Eq.3, (3) vector constraints and (4) symmetries between equations] and mixed integer optimisation [6] can robustly identify the governing equations (1) and (2) up to Re = 10^{10}.
- Further, applying SINDy with open fitting to all 104 possible terms and careful choice of the machine learning parameters, can also recover the correct for of the governing equations (1) and (2), e.g.:-

```
Re: 10^8. Nx, Nz = 1024, 384. Region x:0-1024, z:0-150
Div_x = 55, Div_z = 12, Div_t = 10. 300 subdomains
u_1' = 0.00010069 d^2u_1/dz^2 + -1.00018036 dp/dx + 0.00009931 d^2u_1/dx^2
     + -1.00035329 u_1.du_2/dz + -1.00006506 u_1.du_1/dx
u_2' = 1.00084245 Θ + -1.00082124 dp/dz + 0.00009667 d^2u_2/dz^2
     + 0.00010304 d^2u_2/dx^2 + -1.00084795 u_2.du_2/dz
     + -1.00092491 u_2.du_1/dx
Θ' = 0.00009887 d^2Θ/dz^2 + 0.00010980 d^2Θ/dx^2
     + -1.00107256 Θ.du_2/dz + -1.00132746 Θ.du_1/dx
```

- It should be noted that this open fitting is extremely sensitive to the data selection, the random initialisation and the size and number of subdomains, leading to difficulties in ensuring a converged machine-learning process.

Next step: turbulent transport terms

- Using integrated fluxes derived from the data for the mean and fluctuating components of the Garaud et al. method (i.e. temperature *Θ(z)*, x-velocity $\bar{u}_1(z)$, z-velocity, $\bar{u}_2(z)$, $R(z)$ (= $\Sigma_i R_{ii}$) for i = 1,2, $R_{ij}(z)$ for i,j = 1,2, $F_i(z)$; i = 1,2, $Q(z)$), we will apply the converged machine learning method for obtaining the governing equations to recovering values for the Garaud coefficients and examining variations across the flow.



References:
[1] Garaud, Ogilvie, Miller & Stellmach, MNRAS **407**, 2451-2467, 2010.
[2] Burns, Vasil, Oishi, Lecoanet & Brown, Physical Rev. Res. **2** (2) 023068, 2020.
[3] http://dedalus-project.org
[4] Brunton, Proctor & Kutz, PNAS **113 (15)**, 3932-3937, 2016.
[5] Ascher, Ruuth, Spiteri, Applied Numerical Mathematics **25** (2-3), 151-167, 1997.
[6] Gurobi Optimization LLC, Gurobi Optimiser Reference Manual 2023; v. 9.1.0

Funded by the European Union

European Research Council
Established by the European Commission

Acknowledgements: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. D5S-DLV-786780). The calculations for this paper were performed on the University of Leeds ARC4 facility, hosted and enabled through the ARC HPC resources and support team at the University of Leeds, to whom we extend our grateful thanks.

UNIVERSITY OF LEEDS