



# Turning the tide on potato viruses using data science and machine learning

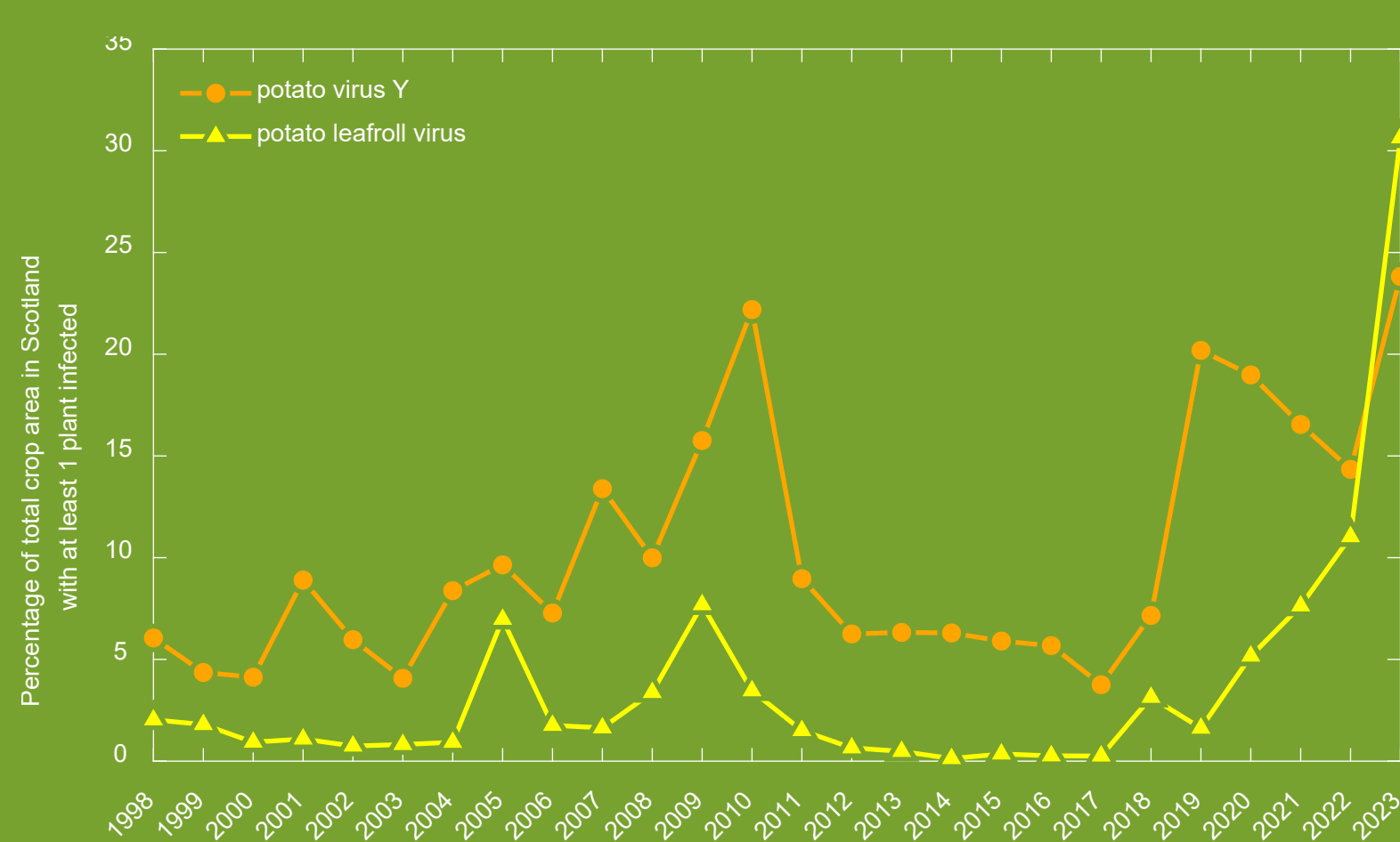
Peter Skelsey  
The James Hutton Institute, Invergowrie, Dundee  
Email: peter.skelsey@hutton.ac.uk



## Introduction

Scottish seed potatoes are a premium global product, and the industry underpins UK potato production that is worth an estimated £4-5bn across all upstream and downstream sectors.

However, incidence of potato-infecting viruses, such as potato virus Y (PVY) and potato leafroll virus (PLRV), has been increasing across Scotland over the last five years, and this is an important concern for the sustainability of the industry:

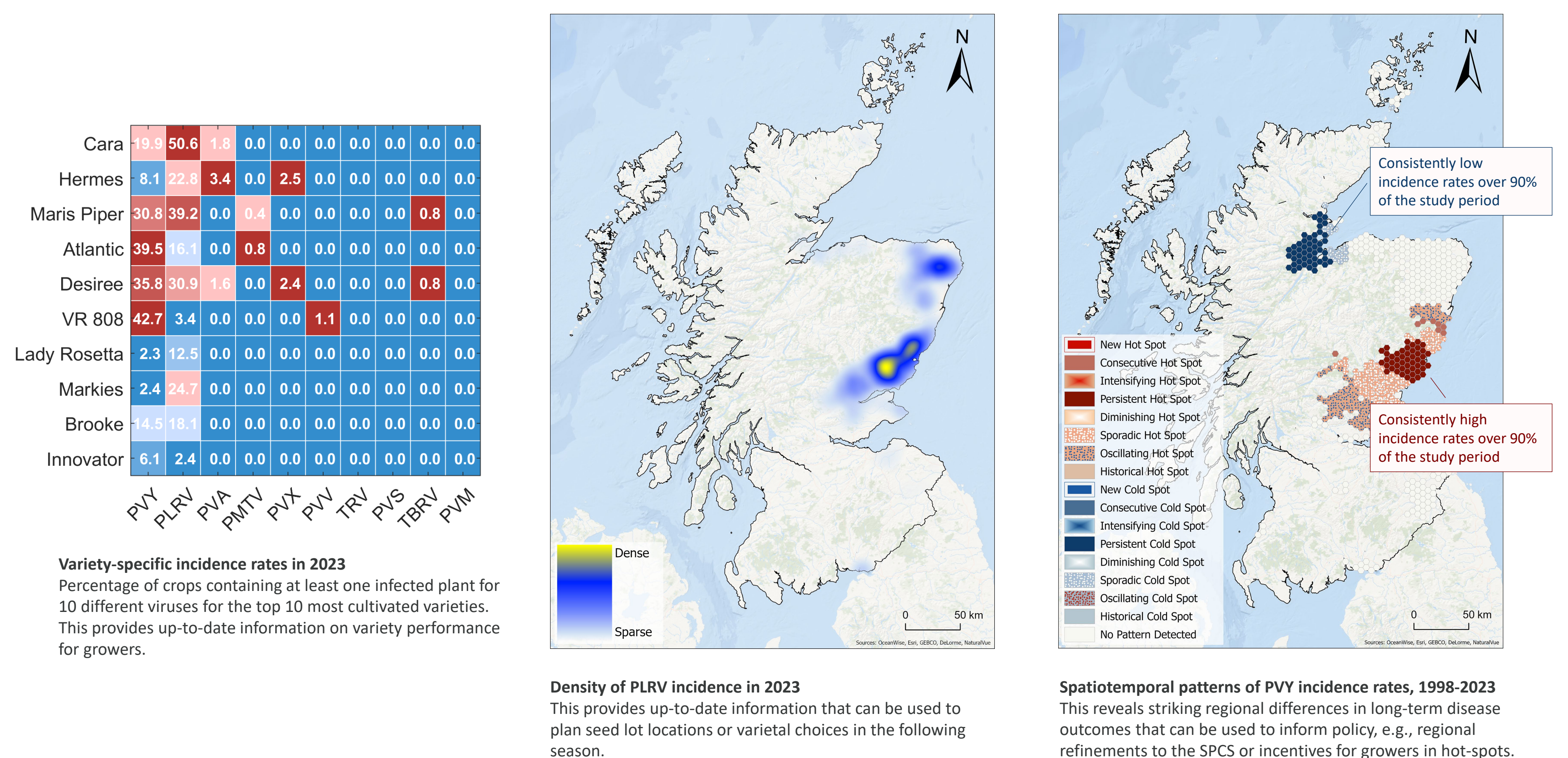


All stages of seed potato production in Scotland are managed and administered under the Seed Potato Classification Scheme (SPCS), which is implemented and regulated by Scottish Government official plant health authorities (SASA), thereby ensuring complete traceability and quality assurance.

Under the SPCS, all growing seed potato crops are inspected 2-3 times a year for a range of faults, and this provides a rich source of data for analysis to inform industry and policy, and for the development of new models to guide decision-making for improved virus management.

## Analysis of data from the Seed Potato Classification Scheme

Data on 10 different potato viruses from over 100,000 crops (1998-2023) were analysed under the RESAS Strategic Research Programme, project JHI-A1-1 (Epidemiology of Key Pests & Diseases). This provided a new overview of the virus situation in Scotland:



## Development of warning systems for forecasting potato viruses

The above results were used to leverage additional funding from the Plant Health Centre to develop national- and local-scale models for forecasting PVY and PLRV, using machine learning applied to the SPCS data and aphid vector data.

**National-scale models:** performance of the best models for forecasting if total virus levels in Scotland in the upcoming season will be lower or higher compared to the average season:

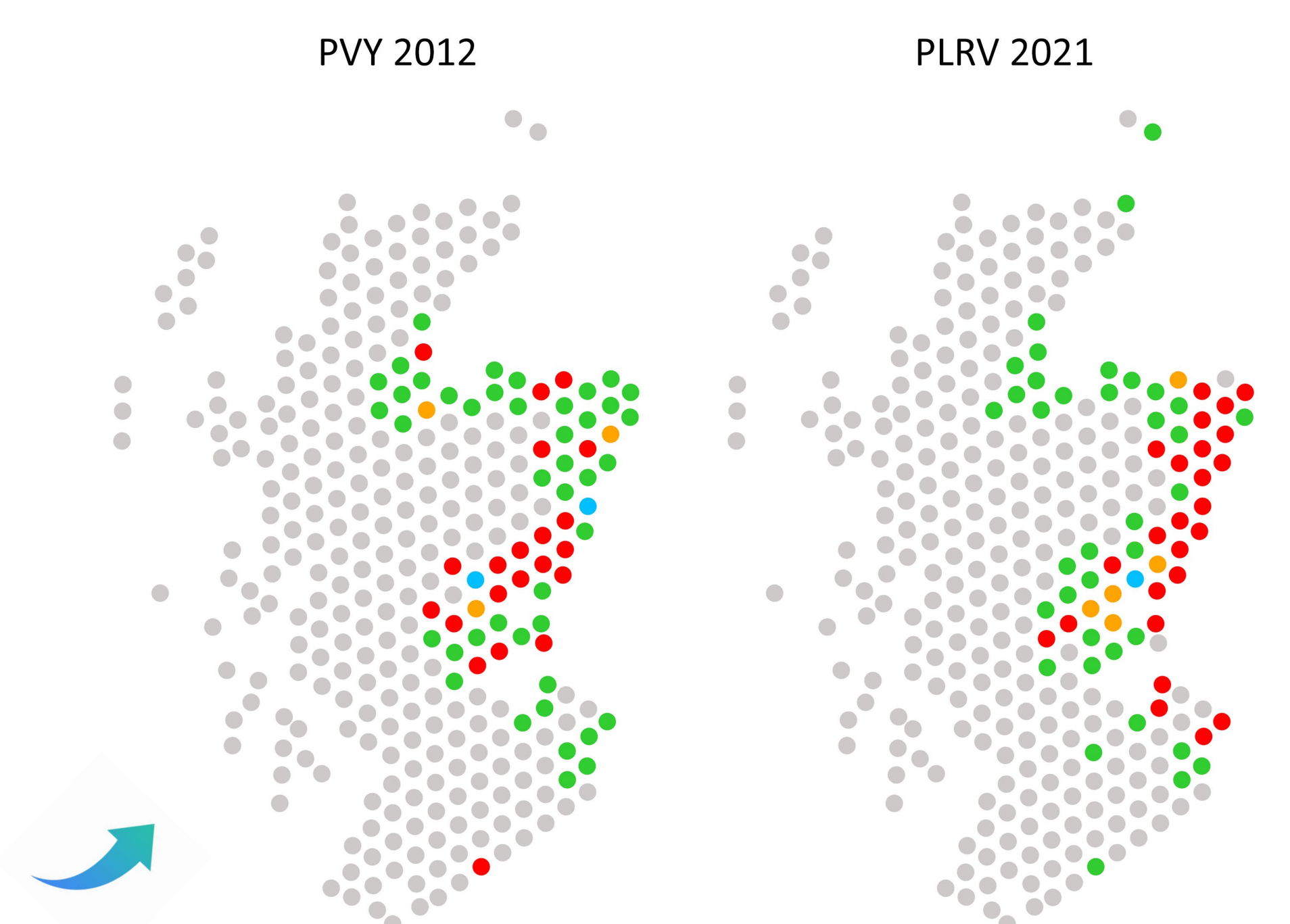
Virus & Algorithm	PPV	NPV	Accuracy
PVY: SVM	1	0.92	96%
PLRV: ANN	0.87	1	92%

PPV = positive predictive value = probability that a prediction of higher virus levels is correct  
NPV = negative predictive value = probability that a prediction of lower virus levels is correct  
SVM = Support Vector Machine  
ANN = Artificial Neural Network

**Local-scale models:** performance of the best models for forecasting if gridded (250 km<sup>2</sup>) virus levels in the upcoming season will be lower or higher compared to the average season for that grid cell:

Virus & Algorithm	PPV	NPV	Accuracy
PVY: RF	0.81	0.94	89%
PLRV: RF	0.86	0.98	94%

PPV = positive predictive value = probability that a prediction of a higher cell value is correct  
NPV = negative predictive value = probability that a prediction of a lower cell value is correct  
RF = Random Forest



**Local-scale gridded predictions of PVY and PLRV for two example years**  
A negative prediction means that virus levels in the upcoming season will be lower compared to the average level for that grid cell, and a positive prediction means that virus levels will be higher. Red = true positive, orange = false positive, green = true negative, blue = false negative.

## Policy implications

- Scottish seed potatoes still have high virus health, but a coordinated effort is required to flush virus out of our production systems.
- This research provides new and comprehensive information on the evolving virus situation in Scotland and the evidence required to develop solutions adapted to the industry's needs.
- The new warning systems for PVY and PLRV will enable data-driven decision-making on virus management that can strengthen the resilience of our seed and ware potato industries.
- The outcomes of this work will be used to update the white paper "Sustainability of Virus Health Management in Scottish Seed and Ware Potatoes" (Scottish Aphid-Borne Virus Research Consortium).

### Acknowledgements

This work was supported by funding from the Rural and Environment Science and Analytical Services (RESAS) Division of the Scottish Government under project JHI-A1-1 (Epidemiology of Key Pests & Diseases), and from the Centre of Expertise for Plant Health under project PHC2023\_08 (Accurate Potato Virus Forecasts). Special thanks go to SASA, FERA, and the Rothamsted Research Insect Survey for data provision.

