Project title:	Precision Agriculture: AI- and Expert-based approach to forecast fruit production in high intra-field variation settings
Project number:	N/A
Project leader:	Jack Stevenson, University of Lincoln
Report:	Annual report, November 2021
Previous report:	N/A
Key staff:	Prof. Stefanos Kollias Dr. Mark Else Dr. David May
Location of project:	University of Lincoln
Industry Representative:	Richard Harnden, Berry Gardens, Unit 20, Wares Farm, Redwall Lane, Linton, Maidstone, Kent ME17 4BA
Date project commenced:	[24 September 2020]

DISCLAIMER

The results and conclusions in this report are based on an investigation conducted over a one-year period. The conditions under which the experiments were carried out and the results have been reported in detail and with accuracy. However, because of the biological nature of the work it must be borne in mind that different circumstances and conditions could produce different results. Therefore, care must be taken with interpretation of the results, especially if they are used as the basis for commercial product recommendations.

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

Jack Stevenson

PhD Student/Lead investigator

University of Lincoln

Signature .... Date ....06/01/2022.....................

[Name]

[Position]

[Organisation]

Signature .... Date ....

Report authorised by:

Stefanos Kollias

Professor of Machine Learning

University of Lincoln

Signature .... Date ....06/01/2022......................

[Name]

[Position]

[Organisation]

Signature .... Date ....

CONTENTS

GROWER SUMMARY

Headline

Merits of quantifying environmental variation to improve forecasted yield estimates and reduce operating costs by forecasting polytunnel phyto-climate using deep learning techniques

Background

Depending on the farm, a polytunnel for strawberries can be configured in a variety of ways, such as the number of tables in the tunnel, the type of ground in the tunnels, the length of the tunnels, the ability to "air" the tunnel via vents and doors, etc. This variation can affect the growing environment and best practices that should be applied to the crop growing in the tunnel, as well as the strawberry cultivars that can be grown in the tunnel. By quantifying this variation, we hope to be able to help growers understand how their tunnel is affecting the optimal growing environment of the strawberry crop.

From this, a strawberry plant's yield can be considered a function of its genetics, environment, and external management. One useful tool for approximating a non-linear function such as yield is neural networks (NNs). These computer programs learn a dataset to solve a given problem, such as image classification or next word prediction. In terms of the yield prediction problem, we would need to consider a variety of factors that can stress a plant, such as disease, irrigation amount, temperature, air flow, pest management, etc. Furthermore, for NNs to provide a meaningful output they need a large amount of data, and when considering a tunnel phyto-climate, there is usually one or two observations for the entire tunnel, which is also used across an irrigation block. Whilst this can provide an idea of the tunnel, it cannot fully describe the environment of the tunnel. This indicates a need to create a more detailed image of the tunnel environment, which can then be fed into a NN for it to learn how the tunnel changes during the growing season. We talk more about an indicative dataset for polytunnel temperature and humidity, and what this can begin to show us in the results section below.

Summary

This project focuses on quantifying how the variation of polytunnel configurations used to grow strawberries affect the temperature and humidity of the polytunnel Phyto-climate, two factors that are important in both strawberry yield and disease management. This should lead to further understanding on how different parts of the tunnel interact, as it can be inferred that different regions grow differently. As can be seen in Table 1.A, the total yield from Riseholme in the 2021 growing season is lower towards the centre of the tunnels, with the west tunnel providing less total yield over the east tunnel.

Table 1.A: Average row yields from the 2021 growing season, along with the row placement within the tunnel.

Row Number	Row placement	Total usable yield
1	Field East Edge, East Tunnel East Edge	4662.557
2	East Tunnel East Centre	4125.160
3	East Tunnel Centre	4280.029
4	East Tunnel West Centre	4678.116
5	East Tunnel West Edge	4084.242
6	West Tunnel East Edge	3415.271
7	West Tunnel East Centre	3504.293
8	West Tunnel Centre	3323.787
9	West Tunnel West Centre	3517.371
10	Field West Edge, West Tunnel West Edge	4121.618

Initial results using the west Riseholme tunnel indicates that there is a greater variation in humidity over temperature, and as a tunnel in the northern hemi-sphere, the south side of the tunnel was warmer than the rest of the tunnel on average. These results were obtained from multiple temperature and humidity sensors placed in the tunnel, providing a higher spatial resolution of data than what is typically gathered for polytunnels.

At the current state in the project, we are purely looking at quantifying the temperature and humidity of a single tunnel. This is due to temperature and humidity affecting both how the plant grows, as well as how disease can grow and spread across the tunnel, if left unmanaged. Often a single observation point is used to represent the entire environmental state of the tunnel. This means that decisions regarding irrigation and humidity control are only representative of the observed area. By using small wireless sensors in a regular grid to create a more accurate representation of the tunnel, we hope to be able to build a better understanding of a polytunnel environment at a given timestamp and how it can affect plant management and fruit production.

Financial Benefits

At the project's current state, we claim that this project can lead to cost savings through a reduced number of crop disease spraying or application of irrigation to the crop such that the crop is not under a water/nutrition related stress. These will lead to concrete numbers as to the financial benefits from the project's work.

Action Points

By the submission of the next annual report, we aim to have some recommendations on how a polytunnel configuration can affect fruit yield.

SCIENCE SECTION

Introduction

The research aims of this project are as follows:

- Build a basic understanding of the patterns in the variation of temperature and humidity in polytunnels.
- Create a deep learning model that can estimate the temperature and humidity of a polytunnel with a 0.5-unit tolerance.
- Generalise this model such that it can work with tunnels of any size and configuration

Experimental work for this project in this year involved monitoring the temperature and humidity of the west tunnel of Riseholme over the months of August and September. This involved using small Bluetooth sensors spread over the tunnel in a grid of size 5 x 9. This allows for a good spatial resolution of data over the 9M by 25M tunnel.

The rest of this section is structured in the following manner: First we describe the farm used and how we set the sensors up to gather the required granularity of temperature and humidity data. Next, we discuss aggregations of the data in both the temporal and spatial domain, to build an understanding of what the dataset is showing about the tunnel configuration's effect on temperature and humidity. Then, we discuss what these results suggest, as well as some of the impact it can have on the crop. Finally, we draw conclusions based on the data, as well as lay out plans for extending the work.

Materials and methods

Key Points:

Main Site Location: University of Lincoln, Riseholme Campus

Soil Type: Coir

Trial Design: Deploy sensors in a grid of where the distances between sensors is 1.5M by 3M, resulting in a 5x9 array

Treatment List: No treatment applied

Assessments: Distances were based on the tunnel setup to be deployed in a regular fashion, to aid the proper processing of data. In addition, observations were recorded in 10-minute intervals. Crop was sprayed Tuesdays 9am for Powdery mildew, and irrigation was based on coir moisture content in 3 locations. Harvest was every Friday between 7am and 11am.

Experimental Process

The experiment was set up at the Strawberry tunnels on the University of Lincoln Riseholme Campus. Strawberry plants were grown in coir substrate and were sprayed on Tuesdays at 9am for Powdery Mildew.

The Bluetooth sensors were deployed with a spatial distance of 1.5M in the East/West direction, 3M in the North/South direction. The east - west distance was determined by the spacing between the tables and the north – south distance was determined by the distance between the table support poles. For the temporal dimension of the data, the sensors recorded an observation of temperature and humidity every 10 minutes

To analyse the data, we have used central tendencies over the temporal dimension to build an understanding on how the data varies in the spatial domain. We use differing timescales, namely daily and weekly as our mean points for temperature and humidity over the rows. For analysing the data in the temporal domain, we take the mean, minimum and maximum of each timestep over the entire tunnel. Individual plots for each quantity in the data can be plotted, however for this report we decided to use aggregate values to demonstrate the variation in the data.

Results

Table 2.A*: Table of mean, median, variance and standard deviation for temperature and humidity over the entire timescale and tunnel.*

Metric	Temperature (°C)	Relative Humidity (%)
Mean	17.02	78.94
Median	16.25	81.4
Variance	14.04	140.07
Standard Deviation	3.75	11.83

The central tendencies used were mean, median, variance and standard deviation. As seen in Table 2.A, the temperature of the tunnel is relatively stable with less than 1 degree difference between the mean and median values. Furthermore, the standard deviation is 3.75 °C, which indicates that the temperature is relatively stable, with approximately 68% of recorded temperatures in the range 13.25 – 20.77 °C. Conversely, the humidity is less stable, as the standard deviation is larger, which indicates that the humidity fluctuates more. This can likely be explained by the warmer days causing the automatic irrigation system at Riseholme to water the crop more often.

Mean analysis of the data over each of the rows

We also take the mean temperature and humidity for each row over the course of a week. We use a weekly timescale as a denser timescale can result in reduced readability of the graph. Complementing the low standard deviation in temperature, the lines representing the temperature for each row are close together and often overlap, shown in Figure 2.A. Also, from Figure 2.A we can see that on average row 6, which is the closest to the centre of the field has the highest temperature.

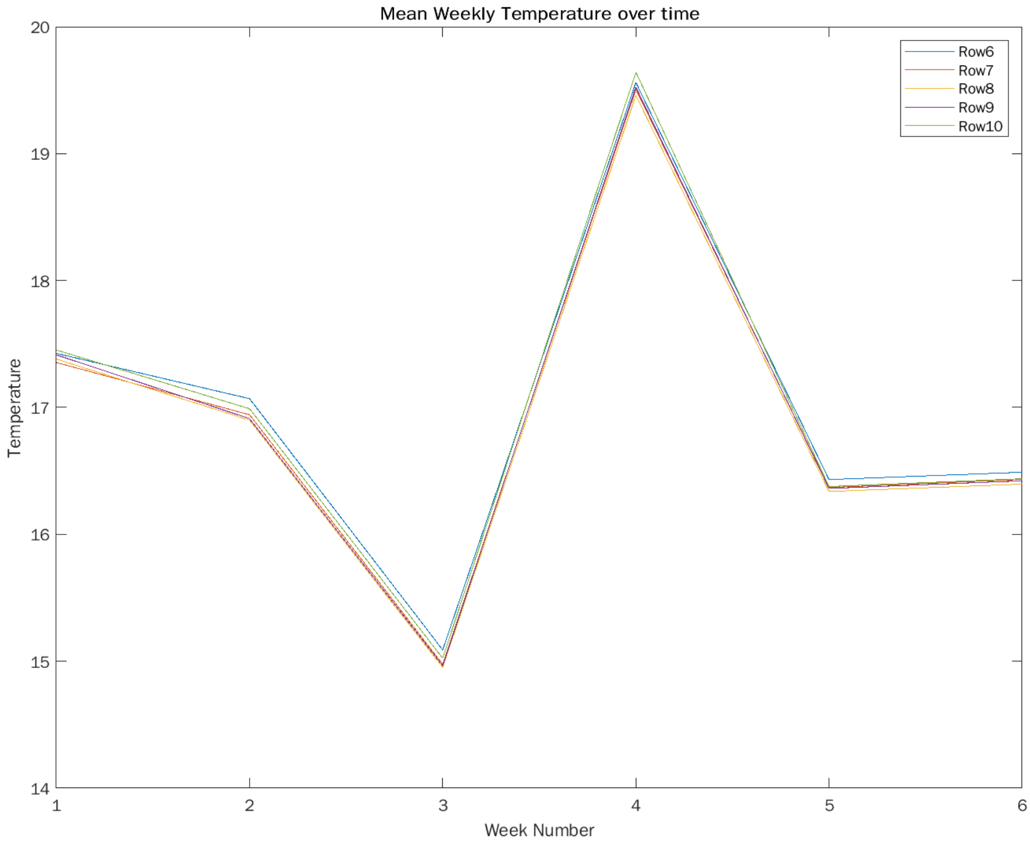

Figure 2.A: Average weekly temperature for each row over the course of the data gathering period.

Furthermore, from Figure 2.B, the average weekly humidity value is higher on the edge rows of the tunnel. In addition, the data also shows a greater variation in the humidity across the rows as each line is more distinct than the temperature for the given row.

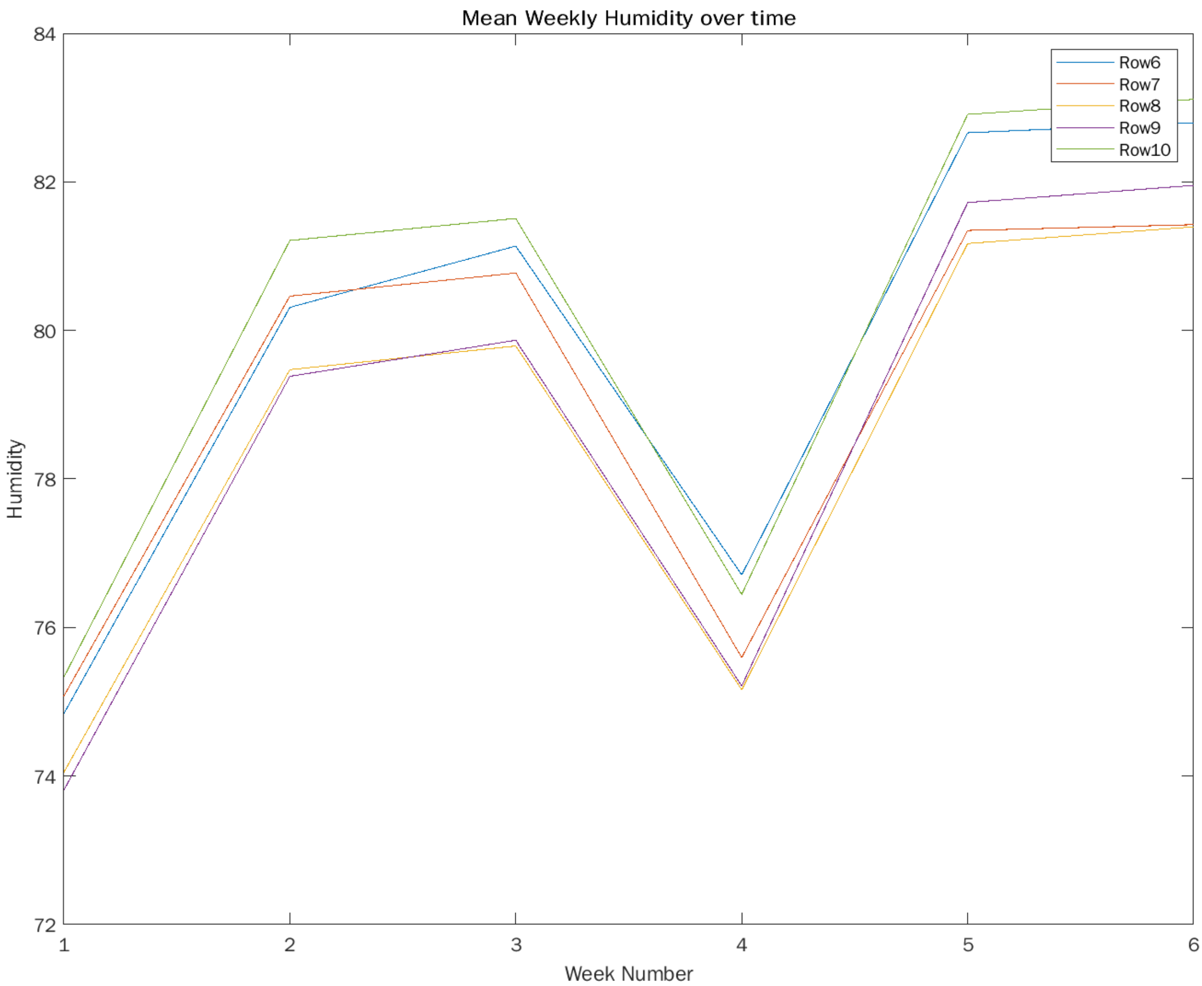

Figure 2.B: Average weekly temperature for each row over the course of the data gathering period.

Spatial Variation in the data

Temperature Variation

The current results have been looking at variation in the temporal dimension, however the gathered data has a spatial component to it as well. We look at the aggregated image of the tunnel, using a 2-dimensional slice of the tunnel, using the data from the bed level of the gathered data. From Figure 2.C, we can see that on average the northern edge of the tunnel is colder, further supported by the minimum and maximum images. Furthermore, the south-eastern most sensor reports the highest observation, with a ~.3 °C increase in temperature over the surrounding area. Conversely, the reading from row 7, column 4 is the coldest observation from the data. Considering the surrounding observations which are close to the central mean temperatures of the dataset, this result is anomalous within the data, which indicates that there is a problem with the sensor.

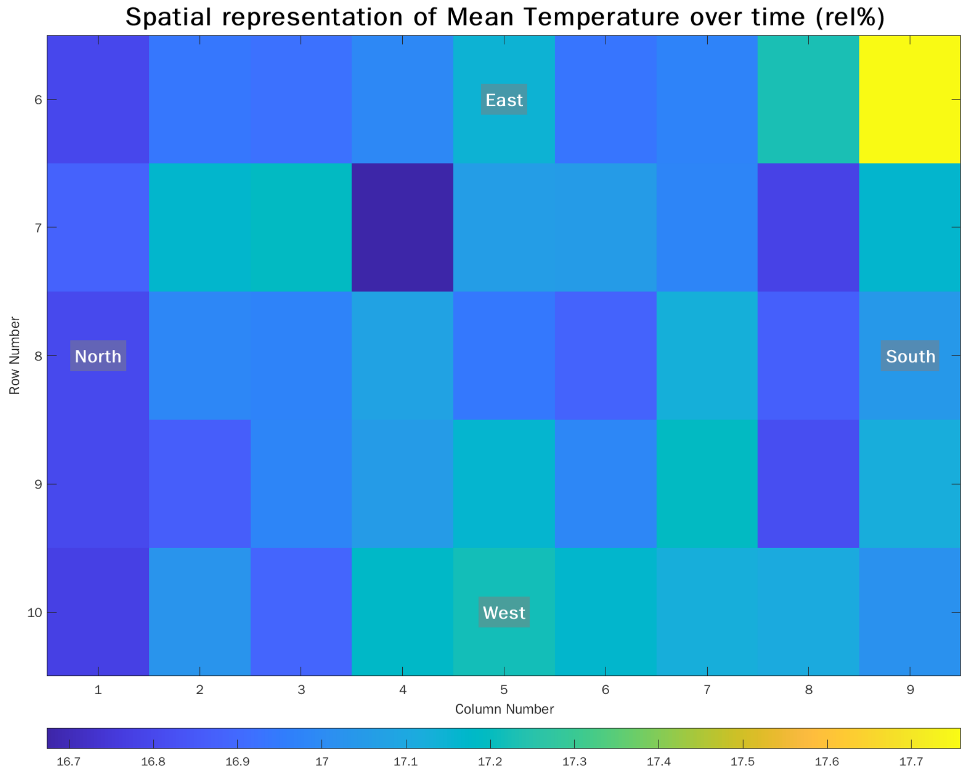

Figure 2.C: Mean Spatial Temperature Representation of the Tunnel.

Looking at the minimum and maximum temperature recorded, (Figures 2.D and 2.E respectively) temperature favours different sides of the tunnel. In the plot of the minimum values, the coldest reported columns are at the north and south sides of the tunnel, where the doors to the tunnels are located. This suggests that in a north-south tunnel like Riseholme, yield will be reduced from plants near the edge of the tunnel due to environmental reactions of the plant such as thermo-dormancy. In addition, the south and western sides of the tunnel also have the highest temperature recorded. The western side recording a greater temperature than the eastern side is likely due to the period the data was gathered in, (August and September), with the sun providing more energy to the western and southern sides of the tunnel. This application of energy, when combined with the south edge reporting colder temperatures indicates that plants closer to the equator, in terms of the individual tunnel experience a greater temperature differential.

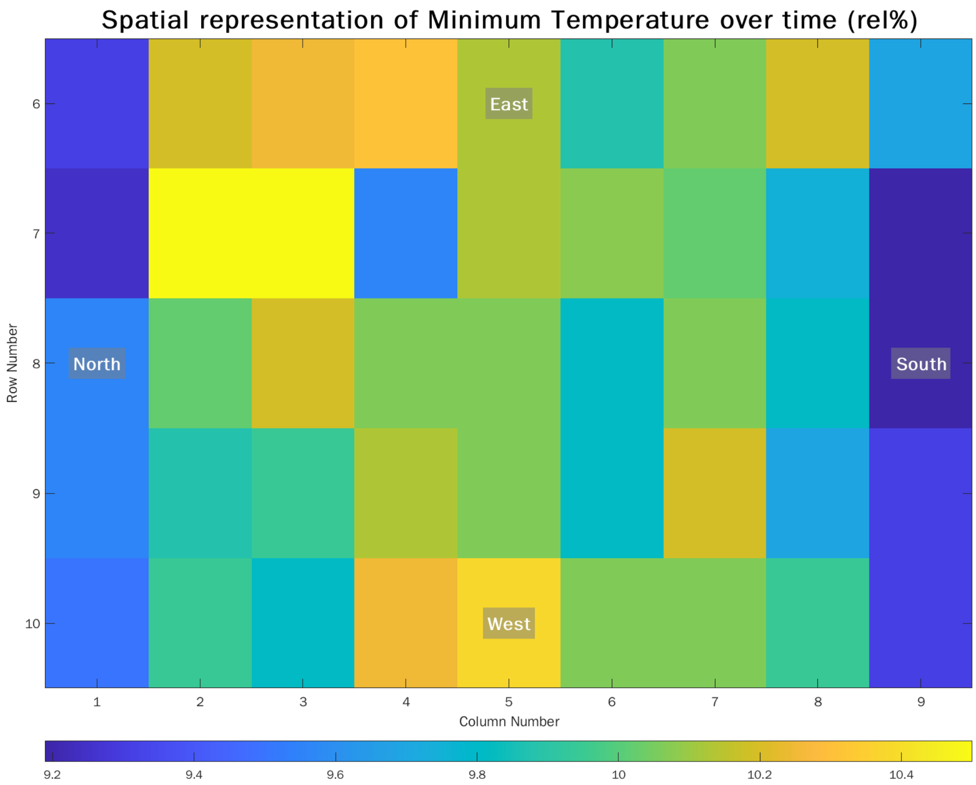

Figure 2.D: Minimum Spatial Temperature Representation of the Tunnel

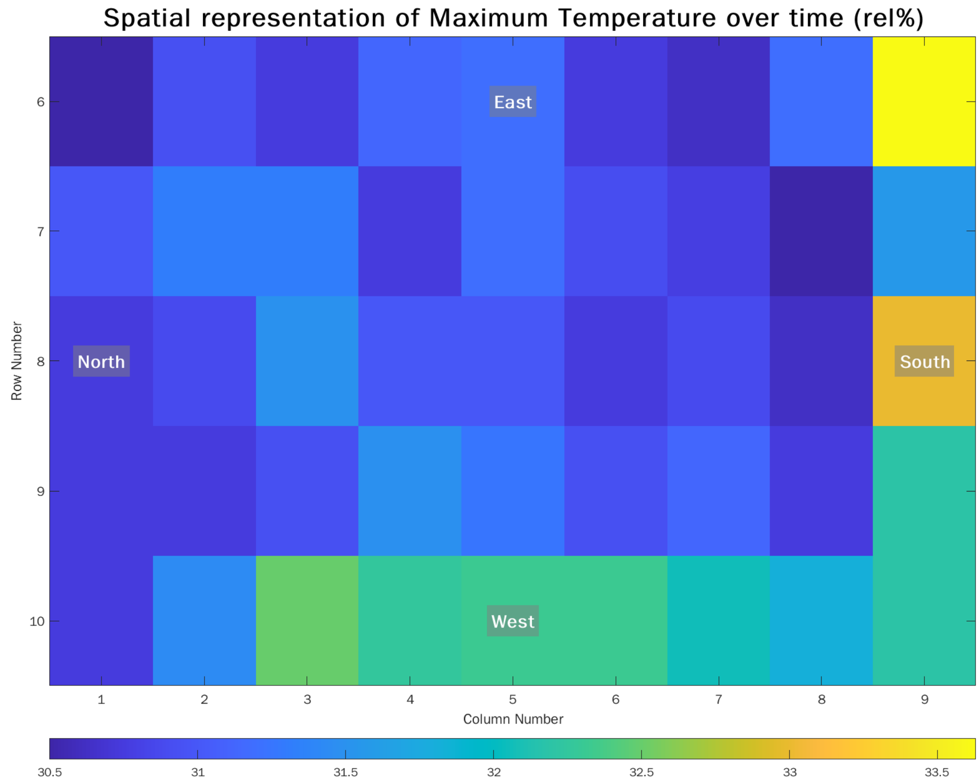

Figure 2.E: Maximum Spatial Temperature Representation of the Tunnel

Humidity Variation

With the mean spatial variation of humidity, seen in Figure 2.F, there is greater activity within the spatial domain, as many results are closer to the mean, as opposed to the minimum in temperature. The anomalous reading for humidity can be seen in row 7 column 2, which has a much lower reading than the other sensors. This suggests that there might be a problem with the sensor device, like the sensor at row 7 column 4 for temperature. In addition, the sensor in row 6 column 9 is also reporting the highest values for humidity, which combines with results from the temperature variation suggests that the calibration for that sensor is off by a significant margin.

Unlike temperature, which has a clear favourable side over time, the humidity is more normally distributed. This suggests that the relative humidity within the tunnel does not have a strong correlation with temperature.

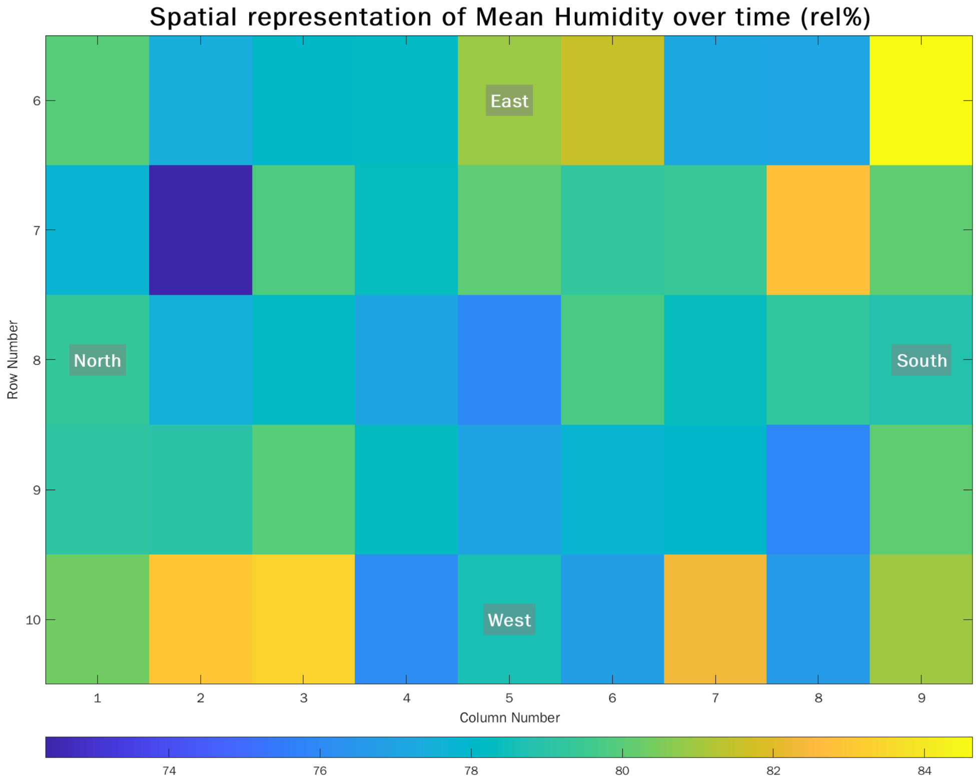

Figure 2.F: Mean Spatial Humidity Representation of the Tunnel.

From the Maximum humidity recorded, many sensors reported a relative humidity of 100%, as seen in Figure 2.G. It can also be seen that the sensor in row 7, column 2 is reporting a lower maximum humidity value, and being the same sensor reporting anomalous observations for the mean humidity, suggests that the sensor was faulty in some manner.

In contrast to the Mean and Maximum representations of humidity, the minimum representation of humidity, Figure 2.H, suggests an even spread of observed values, with the sensor in row 7, column 2 providing a more normal result. Furthermore, approximately 1/3 in from the north-eastern side and 1/3 in from the south-western side show a higher minimum humidity than the rest of the tunnel, which suggests that these areas may be more prone to a higher humidity, such as the wind affecting the observation, or the collection of rainwater not being drained properly from the protective plastic cover.

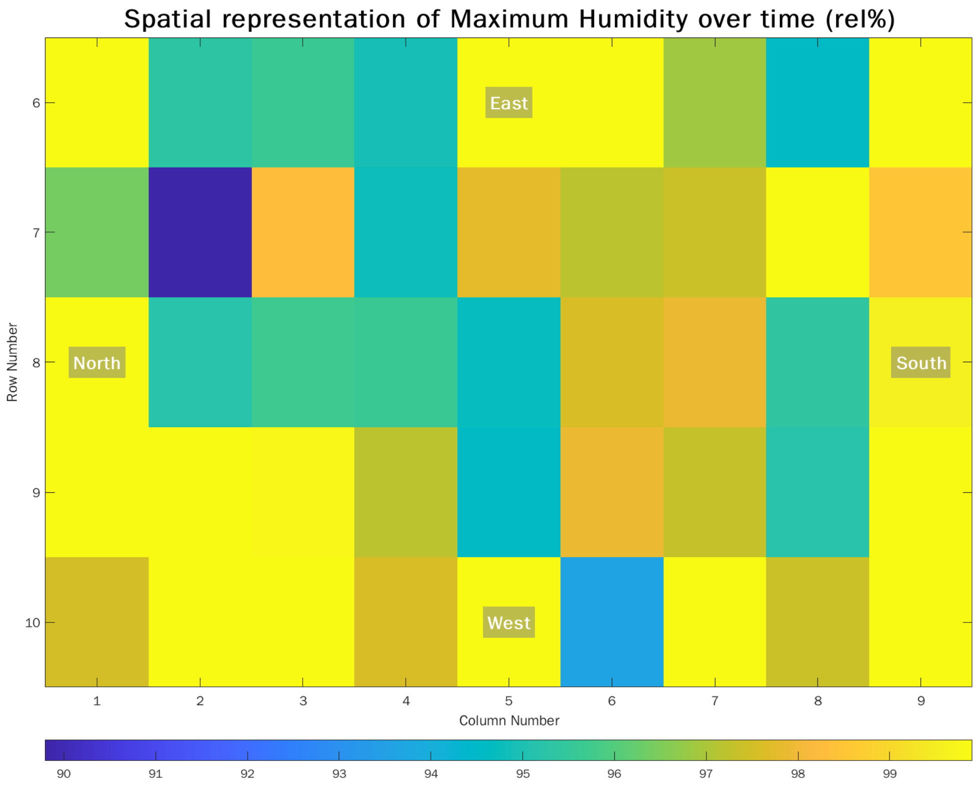

Figure 2.G: Maximum Spatial Humidity Representation of the Tunnel.

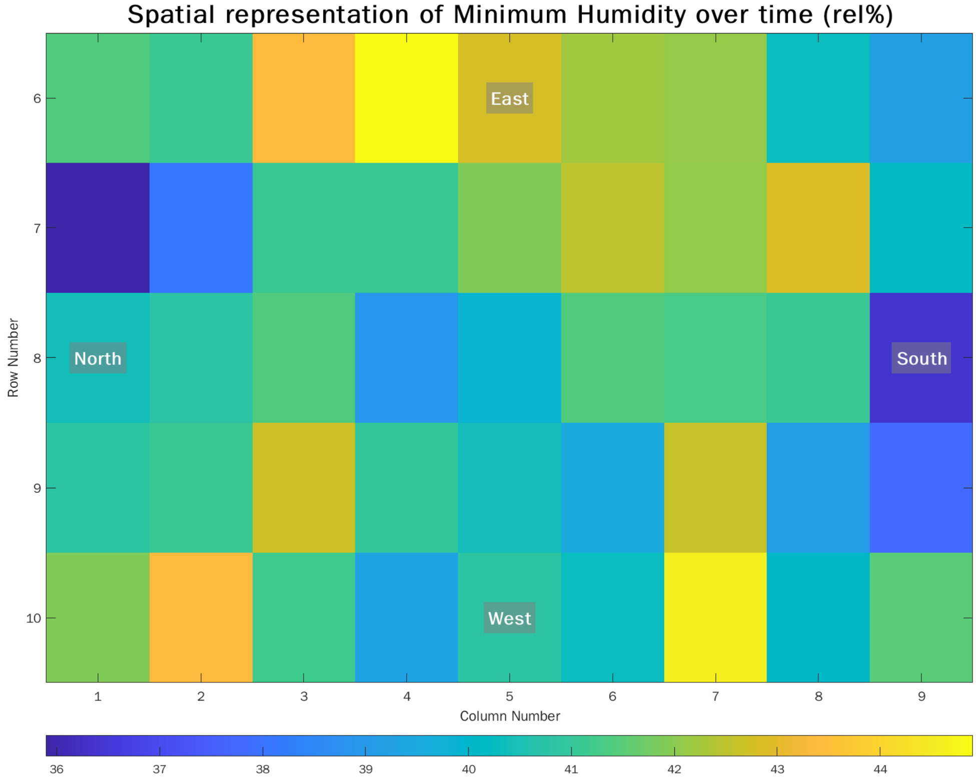

Figure 2.H: Minimum Spatial Humidity Representation of the Tunnel

Finally, it should be noted that these results are not conclusive, as the sensors were located on the eastern side of the tables and were unprotected from external. This can cause the data gathered to be unreliable, however for the purpose of inspecting the variation in temperature and humidity within a polytunnel, it should be considered as appropriate for further investigation.

Discussion

The aggregated spatial changes in the gathered data show that there is variation in the polytunnel phyto-climate, however it is unclear as to the cause of this variation due to the lack of data from being a single tunnel experiment. By increasing the number of tunnels used in the project, we can begin to quantify effects of different tunnel parameters, such as type of ground and age of plastic, on the phyto-climate.

When considering the range of temperatures observed on the south side of the tunnel, it is understood that the change in temperature affects the quality of fruit during the growth period

(Osatuke, A. and Pritts, M., 2021). Although our data was gathered towards the end of the growing season (August and September), Ever bearer plants still grow fruit during this period, and if this temperature differential is observed earlier in the season, then this can also affect the quality of the initial crop, where there is the more fruit produced.

When considering the maximum humidity of the tunnel (Figure 2.G), many of the sensors report a relative humidity value of 100% at some point during the data gathering period. For the sensor located in Row 6, Column 1, this occurred during the night of 14th of September, where the temperature was approximately 14°C (2 sig. fig.).

The weather station for Lincoln, U.K is located at Waddington. On the night in question, it reported values of 14°C and 94% for temperature and humidity respectively (Time and Date, 2021). This suggests that it was a humid night overall, hence the reporting of near 100% humidity for many sensors across the tunnel.

The analysis results from this year are only looking at the surface level of the data gathered. We are currently further analysing the results using machine learning, which can be used to predict a future timestep of the tunnel environment. One well-known class of regression model that are used for this purpose is ARIMA, which uses previously known values to explain a provided time series signal, such as the observations recorded for a single sensor, to build a forecast of the future state (Adhikari, R. and Agrawal, R.K., 2013). An alternative approach is to use a deep learning model, which is trained to use the provided data to infer underlying patterns within the data to make predictions.

Deep learning models are created by stacking multiple specialised layers to complete a given task. Within the forecasting problem space, recurrent layers using units such as the Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) are often used. This is because each unit in the layer is designed to remember the previous state in a time-series, which preserves historical information when generating a prediction. Another Deep learning layer that can be used is the convolution layer, which extracts important information from the previous layer and condenses the information contained within it. By combining these layers, we begin building up a model architecture which can provide a forecast for a given dataset.

Examples of these model architectures are Autoencoders and Transformer Networks. Autoencoders are models that learn how to reconstruct an input by compressing the data to produce a simpler understanding (Goodfellow et al. 2014). One application of autoencoders is in denoising, which will be useful within further analysis as raw sensor data is noisy due to calibration and accuracy errors. Transformer networks (Vaswani, A. et al., 2017) are a type of deep learning model that looks at the global effect on a sequence, by disregarding spatial and temporal aspects of time series data. This allows for a powerful generalisation technique, allowing for the model to be applied to multiple different tunnels. The fact that our data does have a dependency on the positional values in both the spatial and temporal domain needs to be included when running predictive trials with this model architecture.

By analysing and adapting the parameters of these models to affect the learning process, it will enable us to begin understanding how some of the various factors affecting polytunnel design, such as ground type, ground colour, tunnel length and number of tables, influence a plant's ability to produce the highest amount of optimal crop.

Finally, the purpose of this year of work was to identify whether there is a variation effect of the environment of a polytunnel in terms of temperature and humidity. From the results, there is variation, and it can influence the usable yield of the crop, such as row 10, the warmest row in the observed tunnel having a greater usable yield than the other rows. Using the current data as a baseline, future work can be expected to work on generating a more accurate heatmap of the temperature and humidity, increasing the amount of data collected from different tunnel configurations. Furthermore, we will utilise previously mentioned machine and deep learning techniques to provide a greater level of insight into tunnel variation.

Conclusions

Using the data from this year's work, we can conclude the following:

- There is variation occurring within a polytunnel environment, especially with humidity
- Temporal temperature variation across the rows is negligible, however there is merit with spatial temperature variations

Knowledge and Technology Transfer

Project Dissemination 2020-2021

CTP Autumn event 25th November 2020

AHDB Crops PhD Conference 21st – 23rd January 2021

CTP Summer event 1st July 2021

LAR Mini Conference 7th – 9th July 2021

Project Dissemination Plans 2022

- Presentations in conferences/workshops
- Project related Publication in Journal
- Other Industry Events

References

Time and date. (2021) September 2021 Weather in Lincoln. [Website] (Available: https://www.timeanddate.com/weather/uk/lincoln/historic?month=9&year=2021)

Osatuke, A.; Pritts, M. Strawberry Flavor Is Influenced by the Air Temperature Differential during Fruit Development but Not Management Practices. *Agronomy* (**2021)**, *11*, 606. https://doi.org/10.3390/agronomy11030606

Adhikari, R. and Agrawal, R.K., (2013) An introductory study on time series modelling and forecasting. *arXiv preprint*. https://arxiv.org/pdf/1302.6613.pdf

GoodFellow, I., Bengio, Y., and Courville, A. (2014) Deep Learning MIT Press

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

Appendices

Appendix A: Video Visualisations of the full spatial changes in the horizontal plane:

Temperature: https://youtu.be/-78xxUUcswE

Humidity: https://youtu.be/NRbZGrF__k4

Appendix B: 4 Week subset of the gathered data:

https://drive.google.com/file/d/1pAg6WapMjl6IzJkRGnKMd9rsK_4KeYdc/view?usp=sharing

