Supervised learning for data driven cucumber yield prediction and control of greenhouses

Melanie Millet, Andreas Kloos, Lucas Luttner, Fabian Nickel, Meike Haese

1. Data Preparation

- **Deleting missing values:** Total of 11
- **Deleting outliers**: Negative values in irrigation
- **Deleting cycles** → deliver no relevant information
- One-Hot-Encoding country data

cycle	**country**	irrigation	sun	temperature	yield

Envidium_NED	**Glyfanac_ESP**	irrigation	sun	temperature	yield

- Divide dataset into **target values y** („yield“) and **features X**
- Split Dataset into **train**, **validation** and **test data** (test_size=0.25)
- Use min-max **normalization** [0,1] to reduce variance error of NN

2. Neural Network – Process of Optimization

Goal: Train a model that minimizes the loss function

Methods used for selecting model architecture

- **Random-Search:** Train the model with a randomly given combination of hyperparameters that are in a given interval
- **Grid-Search:** Try all possible combinations of the given interval

Specifications

- Small number of Hiddenlayers ➔ small number of features
- choose the model with the smallest MAPE
  - → Percent easy to interpret
  - → Good comparison to models with another context
- Activation function: only ReLU makes sense. Sigmoid limits the prediction at 1

2. Neural Network – Random Search

- Early Stopping – Patience = 30

Parameter Intervals		Best Parameters
Hidden Layer	[2, 4]	2
Units/Neurons	[32, 128]	35,35
Activation Hiddenlayer	[ReLU, Sigmoid]	ReLU
L2-Regularization	[0.05 , 0.0005]	0.02647
Dropout	[0, 0.2]	0.00753
Epochs	500	500
Batch Size	128	128
Optimizer	Adam	Adam
Learning Rate	[0.0002, 0.02]	0.0192

- Net should predict a number in the interval [0,∞] → ReLU

→ **Best MAPE:** 1.509

2. Neural Network – Grid Search

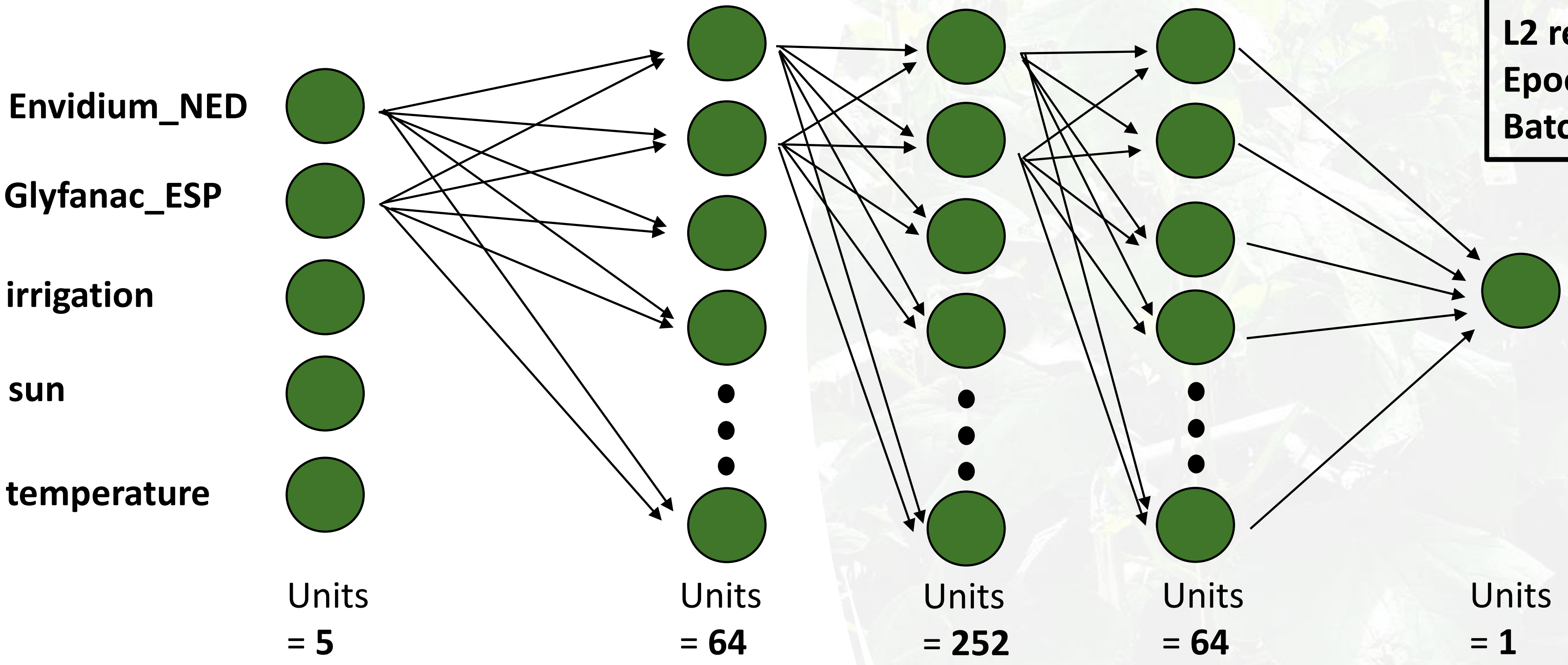
- Using **GridSearchCV** for hyperparametersearch
- 5 Fold Cross Validation (standard)
- Changing parameters with a multidimensional grid

Best mean score: MAPE of 1.48

Parameters							
Hidden Layer	1	2	3				
Units/Neurons	8	16	32	64	150	252	504
Activation	ReLU	Sigmoid					
L2-Regularization	0	0.005	0.01				
Dropout	0.001	0.002					
Epochs	300	400	600				
Batch Size	32	64					
Optimizer	Adam	SGD					
Learning Rate	0.01	0.001	0.0045				

2. Best Neural Network (Evaluation)

Learning Rate: 0.001
Dropout: 0
Activation: ReLU
L2 regularizer: 0
Epochs: 600
Batch: 32



MAPE = **1,30**
→ model does not overfit

3. Cost Function

Setup:

- Initialize a dataframe with all possible combinations
- Only test temperature/irrigation values within boundaries which are known to the net

Add cost calculation:

Cost per cycle	Used formula
Temperature cost	$\lvert test\ temperature - actual\ temperature \rvert * 3600\ EUR$
Irrigation cost	$test\ irrigation \frac{l}{m^2 d} * 60 \frac{d}{cycle} * greenhouse\ m^2 * \frac{1\ m^3}{1000\ l} * 0.021 \frac{EUR}{m^3}$
Labour/depreciation cost	$greenhouse\ m^2 * \frac{20\ EUR}{m^2 * year} * \frac{1\ year}{6\ cycles}$
Cost for underfulfilling demand	$\max((demand - yield), 0)$

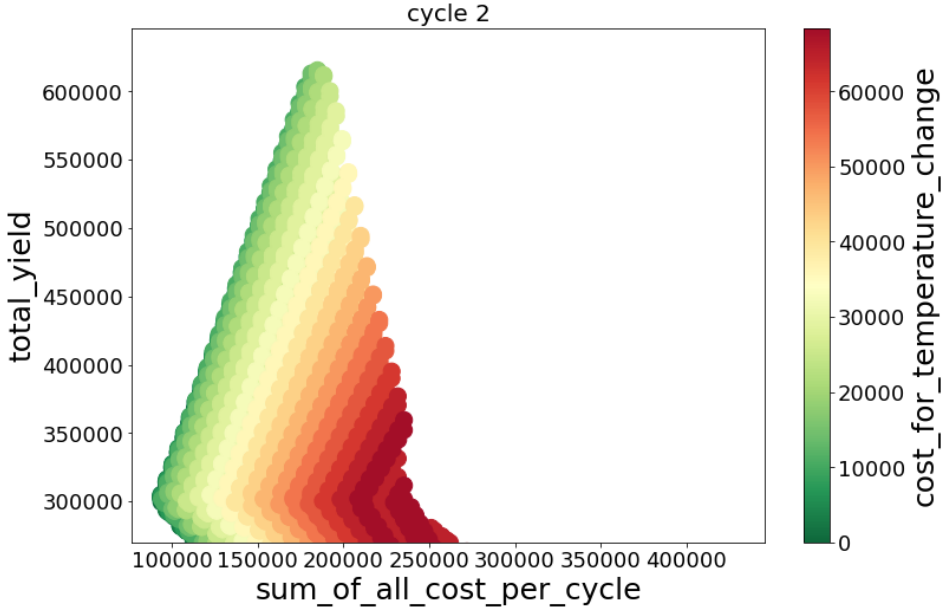
➡ Find minimal sum of costs within dataframe

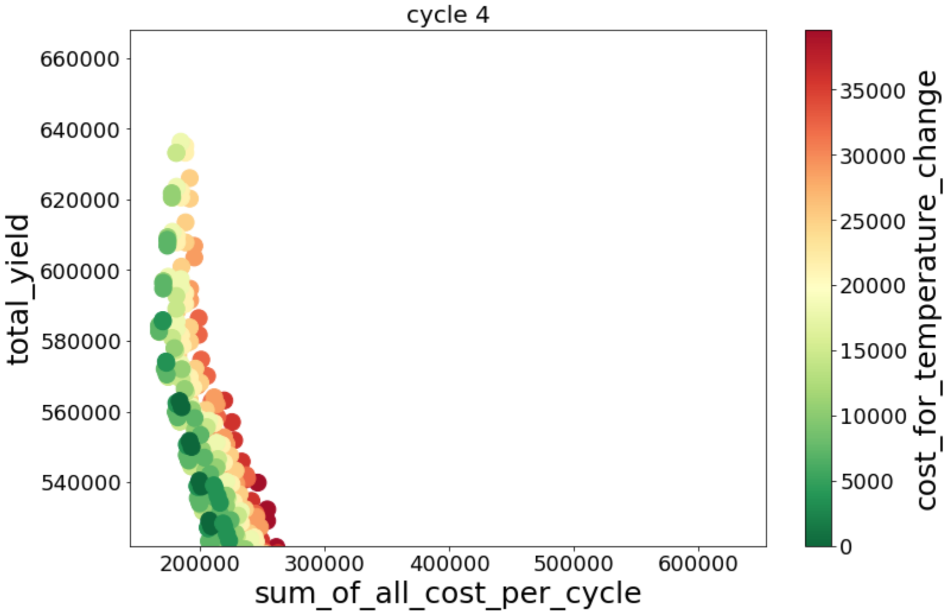
3. Cost Function – Optimal Decisions

Cycle	size	temperature	irrigation	pesticide	demand	Cost for irrigation	Cost for temperature change	Cost for labour/ depreciation	Cost for underfulfilling demand	Sum of all cost per cycle
2	28,000	15	3.5	Gly	300,000	123.48	0	93,333	0	93,457
3	40,000	22	3.5	Gly	500,000	178.40	0	133,333	1,794	136,292
4	48,000	23	3.5	Gly	580,000	211.68	7,200	160,000	0	167,411

For cycle 3 underfulfillment of demand:

- Why not irrigate more? → perfect level of irrigation already reached
- Why not change temperature? → additional cost of min 3,600 EUR not worth it (> 1,794 EUR)
- Why not increase greenhouse size? → additional labour cost of min 3,333 EUR not worth it





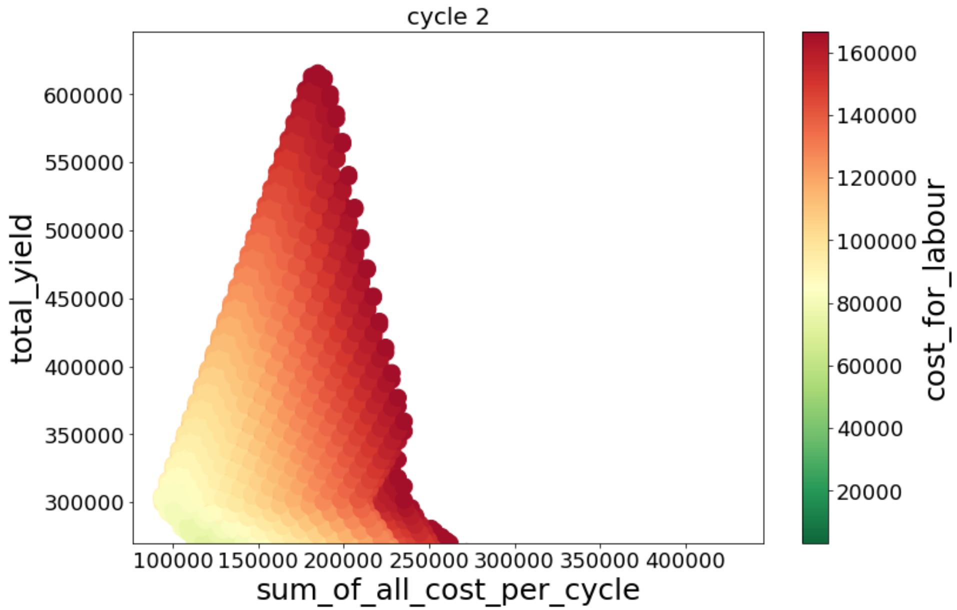
TEMPERATURE COST

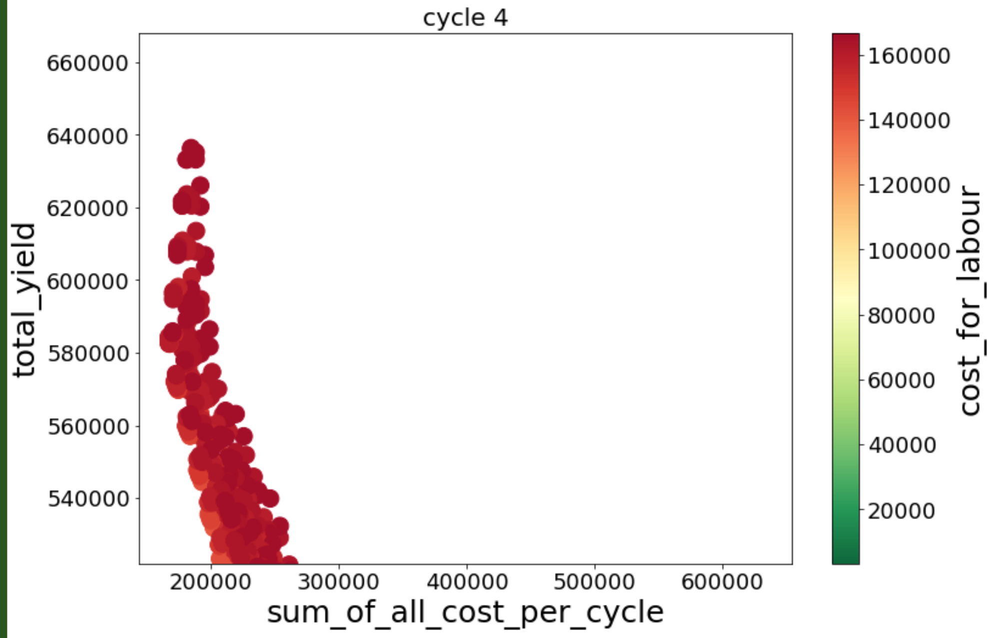
What‘s more efficient?

→ There is no clear winner

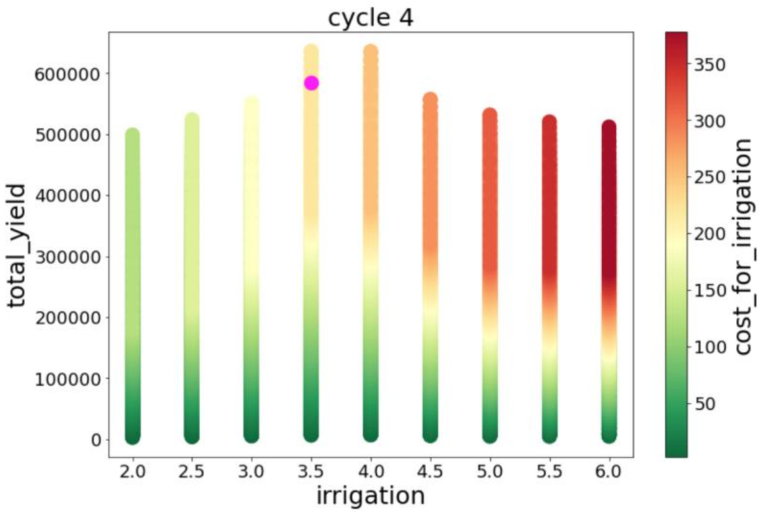
→ Depends on cycle properties (e.g. outside temperature) and demand

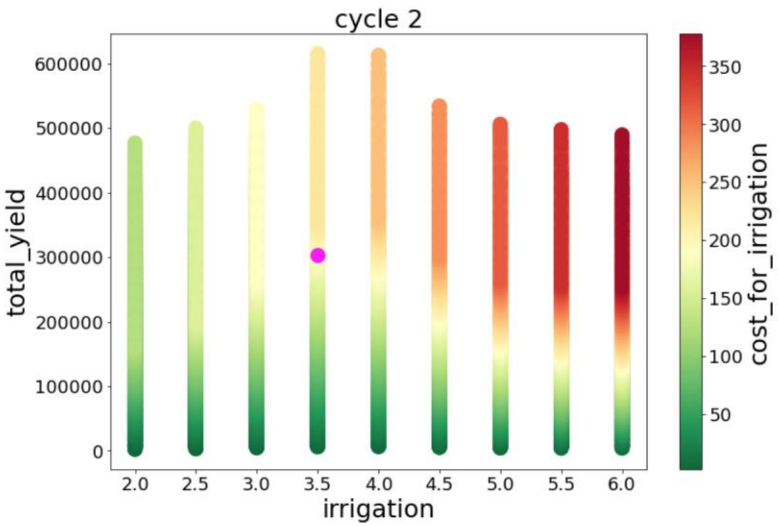
LABOUR COST

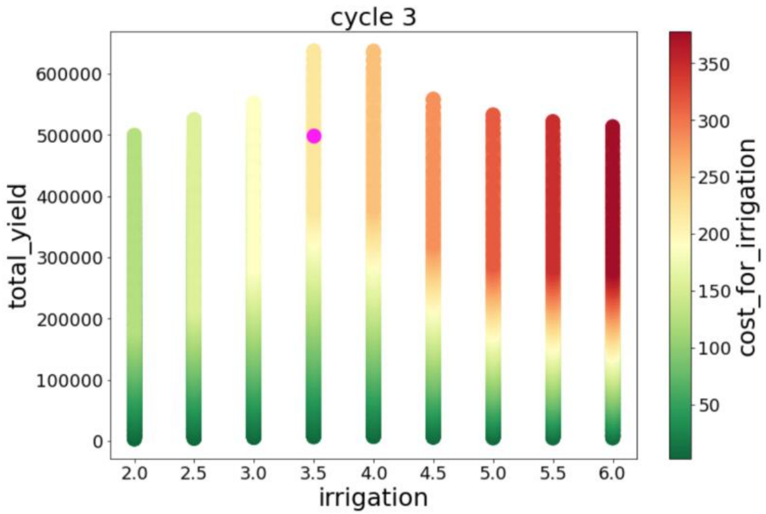




3. Visualization – Cost for Irrigation



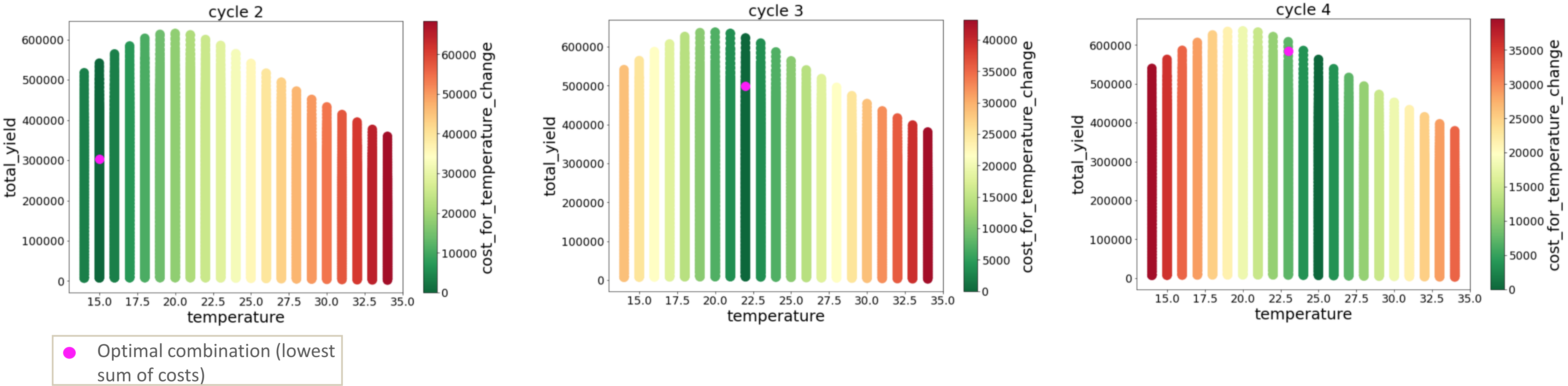




● Optimal combination (lowest sum of costs)

- Regardless of cycle, irrigation is always 3.5
- Reason: cheapest method to increase yield

3. Visualization – Cost for Temperature



- Optimal temperature for cucumber yield is 20°C
- But it‘s never used because of high costs

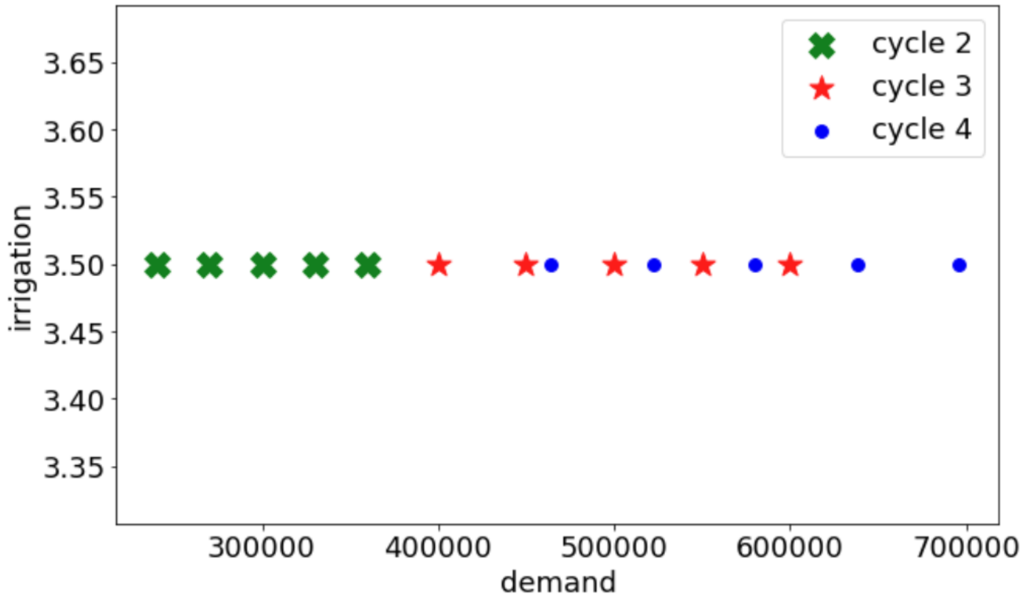
4. Sensitivity Analysis – Optimal Decision and Costs

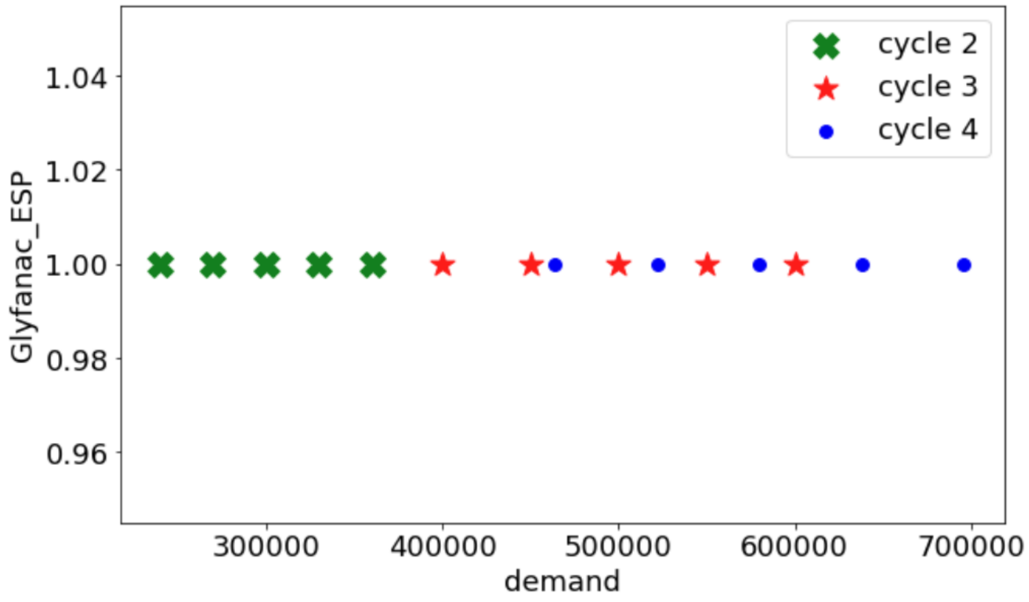
- Conducted to see how sensitive results are with respect to the contracted demand
- Contracted demand is varied in steps of 10% from -20% to +20%, for all three cycles

Optimal decisions and costs for scenarios {0.8, 0.9, 1.0, 1.1, 1.2} = factors for expected demand :

	irrigation	sun	temperature	size	total_yield	Envidum_NED	Glyfanac_ESP	demand	scenario	cost_for_underfulfilling_demand	cost_for_temperature_change	cost_for_irrigation	cost_for_labour	sum_of_all_cost_per_cycle	cycle
0	3.5	0.303956	15	22000	238304.309845	0	1	240000	0.8	1695.690155	0.0	97.02	73333.333333	75126.043488	2
3	3.5	0.303956	15	25000	270800.352097	0	1	270000	0.9	0.000000	0.0	110.25	83333.333333	83443.583333	2
6	3.5	0.303956	16	29000	327451.898575	0	1	330000	1.1	2548.101425	3600.0	127.89	96666.666667	102942.658092	2
9	3.5	0.303956	16	32000	361326.232910	0	1	360000	1.2	0.000000	3600.0	141.12	106666.666667	110407.786667	2
12	3.5	0.303956	15	28000	303296.394348	0	1	300000	1	0.000000	0.0	123.48	93333.333333	93456.813333	2
1	3.5	0.537537	22	32000	398564.575195	0	1	400000	0.8	1435.424805	0.0	141.12	106666.666667	108243.211471	3
4	3.5	0.537537	22	36000	448385.147095	0	1	450000	0.9	1614.852905	0.0	158.76	120000.000000	121773.612905	3
7	3.5	0.537537	22	44000	548026.290894	0	1	550000	1.1	1973.709106	0.0	194.04	146666.666667	148834.415773	3
10	3.5	0.537537	22	48000	597846.862793	0	1	600000	1.2	2153.137207	0.0	211.68	160000.000000	162364.817207	3
13	3.5	0.537537	22	40000	498205.718994	0	1	500000	1	1794.281006	0.0	176.40	133333.333333	135304.014339	3
2	3.5	0.523284	23	38000	462510.339737	0	1	464000	0.8	1489.660263	7200.0	167.58	126666.666667	135523.906930	4
5	3.5	0.523284	23	43000	523366.963387	0	1	522000	0.9	0.000000	7200.0	189.63	143333.333333	150722.963333	4
8	3.5	0.523284	21	50000	633151.531219	0	1	638000	1.1	4848.468781	14400.0	220.50	166666.666667	186135.635447	4
11	3.5	0.523284	21	50000	633151.531219	0	1	696000	1.2	62848.468781	14400.0	220.50	166666.666667	244135.635447	4
14	3.5	0.523284	23	48000	584223.587036	0	1	580000	1	0.000000	7200.0	211.68	160000.000000	167411.680000	4

4. Sensitivity Analysis

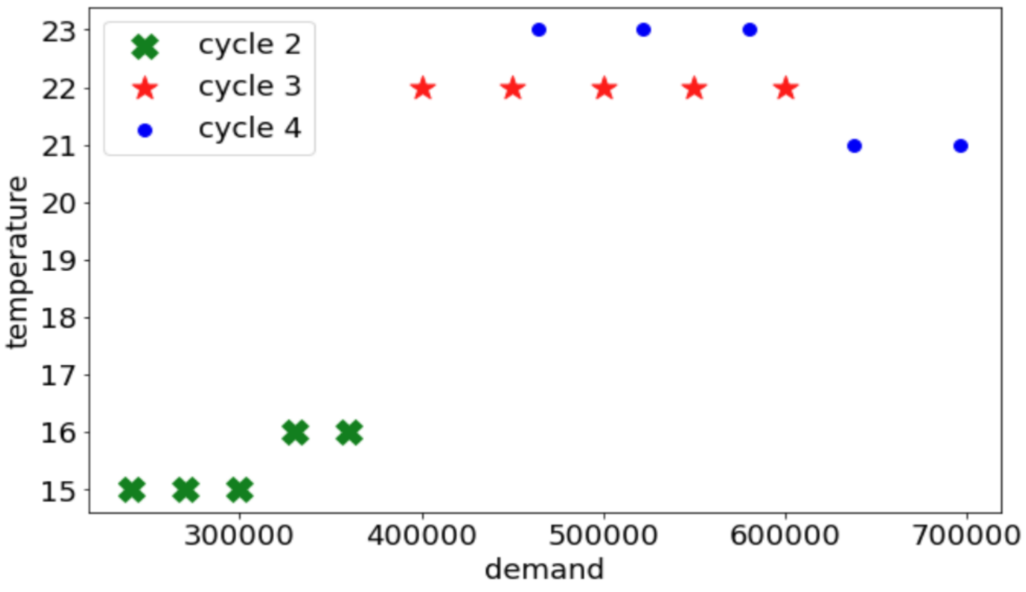


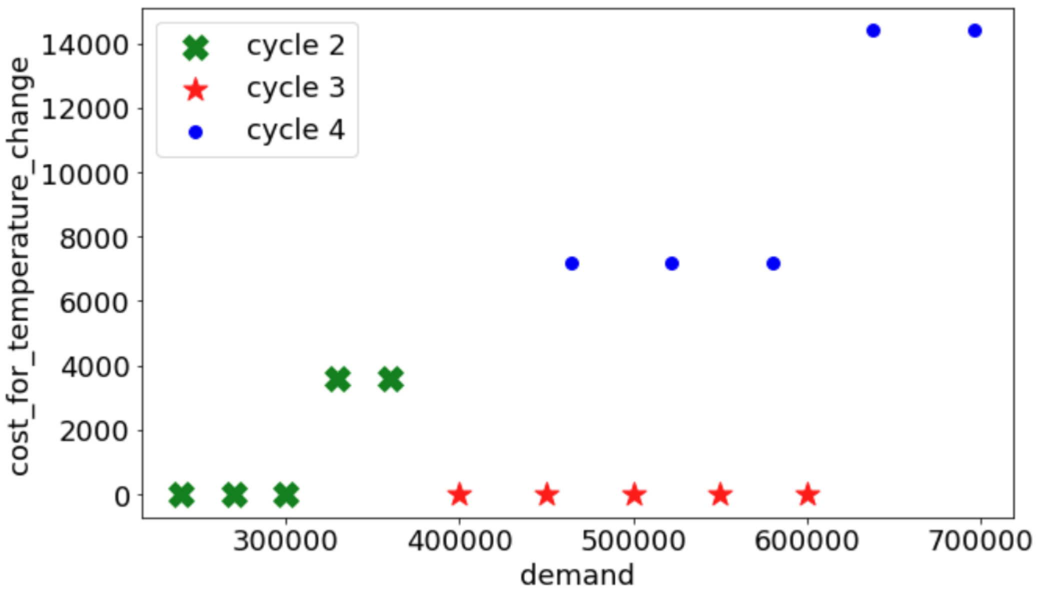


- Ideal irrigation at 3.5 $\frac{l}{sqm*d}$

- For all scenarios and cycles
  Glyfanac was chosen as optimal pesticide

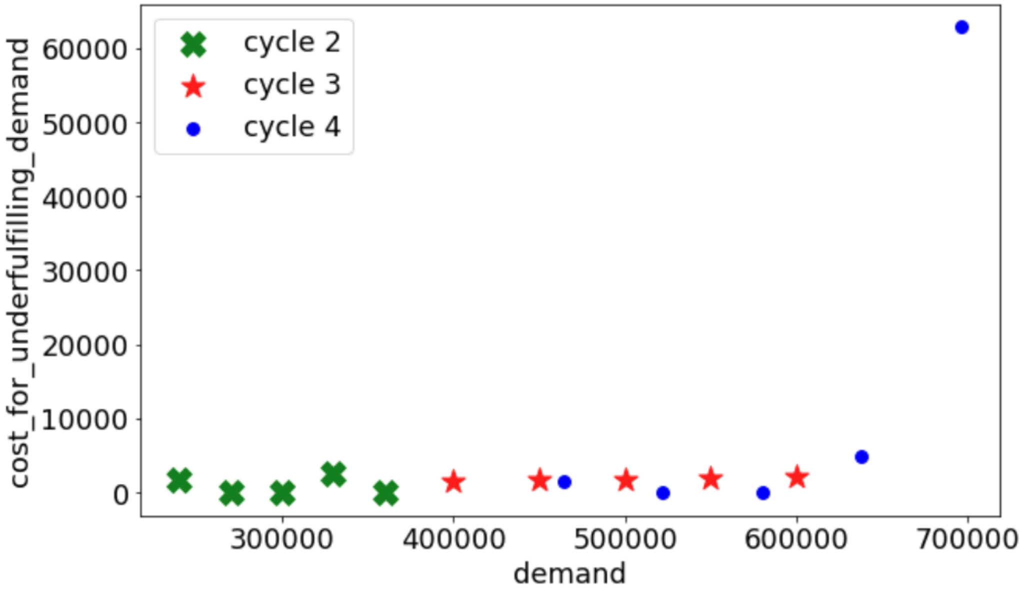
4. Sensitivity Analysis

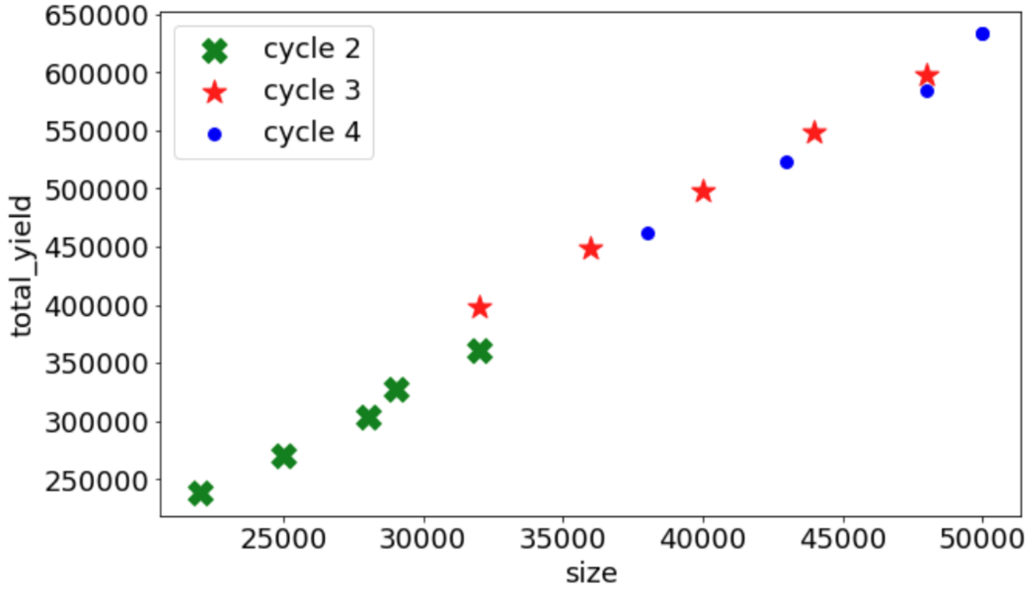




- For cycle 2 temperature changes are only necessary for scenario 1.1 and 1.2
- For each scenario of cycle 3 there are no costs for temperature change, since outside temperature is closer to the optimum
- Cycle 4 has higher temperature costs since hot outside temperatures need to be cooled down, to adress high demand

4. Sensitivity Analysis





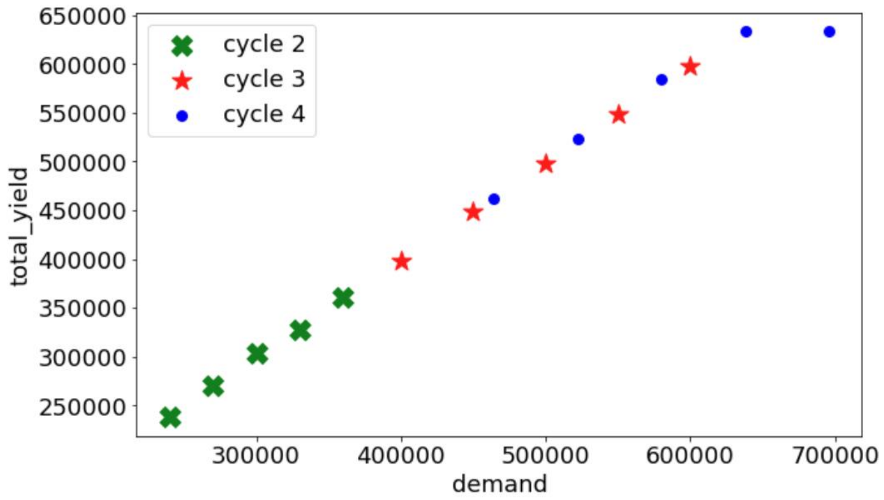
- For most scenarios it is better to underfulfill rather than exceeding the demand
- Only for some scenarios demand is fulfilled
- Limited field size results in increasing difficultiy to supply enough cucumbers → high underfulfilling costs

- To generate a higher yield, larger fields are needed
- Size of fields are limited at 50,000 sqm
- Total yield depends highly on the size of the field and therefore caps at certain point

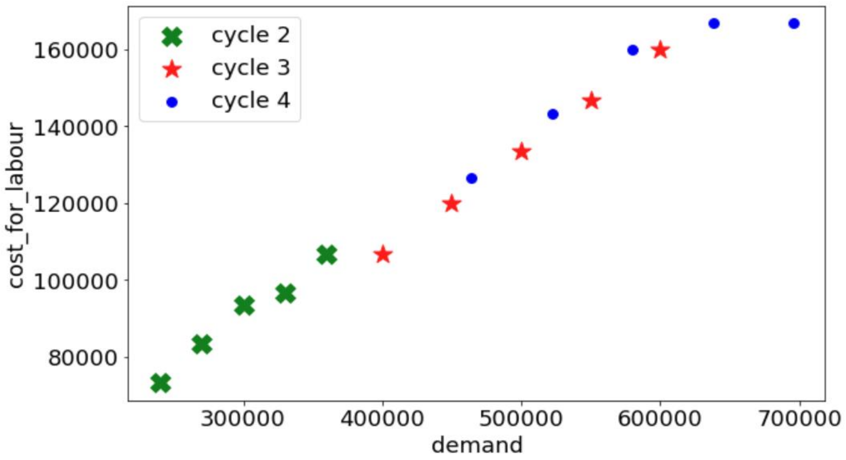
5. Concluding Remarks

- Irrigation is the cheapest method to increase yield → always irrigate optimal at 3.5 $\frac{l}{sqm*d}$
- There are scenarios in which max. capacity is reached (e.g.+20% demand cycle 4)→ resulting in high cost for underfulfilling demand
- Using Glyphanac shows higher yields than Envidum
- temperature change is rather useful at higher demand levels (depends on cycle)
- accepting costs for not meeting demand is an option

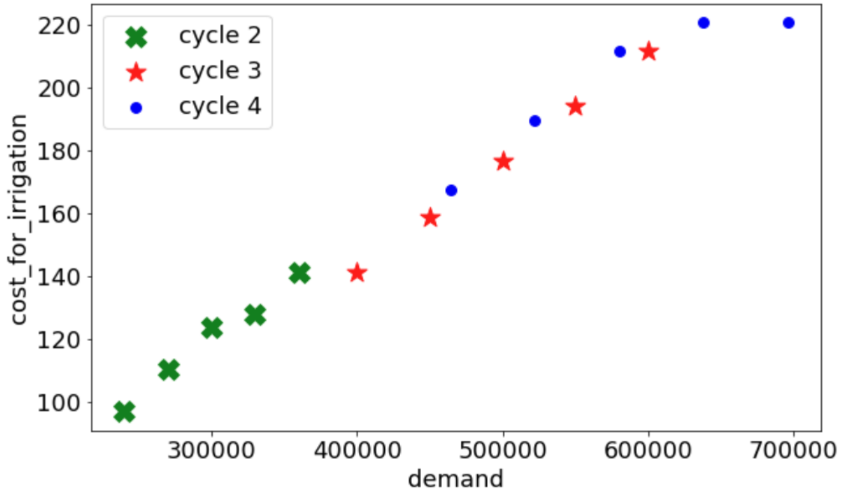
Attachment



650000
600000
550000
500000
450000
400000
350000
300000
250000
total_yield
cycle 2
cycle 3
cycle 4
300000
400000
500000
600000
700000
demand



160000
140000
120000
100000
80000
cost_for_labour
cycle 2
cycle 3
cycle 4
300000
400000
500000
600000
700000
demand



220
200
180
160
140
120
100
cost_for_irrigation
cycle 2
cycle 3
cycle 4
300000
400000
500000
600000
700000
demand


Thank you for
your attention!