

Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in  
Statistica per l'Economia e l'Impresa



RELAZIONE FINALE

**AGRIVOLTAICO:  
ENERGIA RINNOVABILE FOTOVOLTAICA APPLICATA A  
COLTIVAZIONI AGRICOLE.  
ANALISI STATISTICA SULLE CARATTERISTICHE DELLE PIANTINE  
ALLOCATE IN CAMPI CHE DIFFERISCONO PER CONDIZIONI DI  
LUCE VARIABILI**

Relatore Prof. Erlis Ruli  
Dipartimento di Scienze Statistiche

Laureanda: Giorgia Battiston  
Matricola N. 1175032

Anno Accademico 2022/2023



## **ABSTRACT**

La tesi mira a un'analisi che ha come obiettivo l'identificazione di differenze significative della crescita e dello sviluppo di specifiche caratteristiche delle piantine situate in campi diversi, ognuno con condizione di luce propria generata dall'assenza o dalla presenza di un determinato impianto agrivoltaico.

In primo luogo viene presentato l'oggetto di studio ossia l'agrivoltaico, un impianto fotovoltaico applicato a coltivazioni agricole che vira dal comune termine fotovoltaico in quanto viene installato direttamente sopra ai campi di coltivazione e non a fianco com'è solito pensare. L'esperimento si è svolto in campi di diverse coltivazioni tra cui fagioli, mais, patate e erba medica; tra queste si è scelto di analizzare i dati delle coltivazioni di patate.

Le analisi che presidiano l'intero elaborato vedono interessati modelli di regressione polinomiale e modelli additivi generalizzati confrontati fra loro, l'analisi della varianza e infine l'analisi delle componenti principali; si ricorre all'ausilio del software statistico R Studio.



# INDICE

<b>1. INTRODUZIONE</b>	1
1.1 L'impianto fotovoltaico	1
1.1.1 <i>Invenzione del fotovoltaico</i>	1
1.1.2 <i>Composizione e funzionamento dell'impianto fotovoltaico</i>	2
1.1.3 <i>Sustainable Development Goals</i>	3
1.2 Oggetto di studio	4
1.3 Raccolta dei dati	5
1.4 Obiettivi dell'analisi	7
<b>2. IL DATASET E LE VARIABILI</b>	9
2.1 Potatoes	9
2.2 Meteo	10
<b>3. ANALISI ESPLORATIVE</b>	11
3.1 Condizioni di luce confrontate tra loro (Analisi Univariata)	11
3.1.1 <i>Altezza</i>	11
3.1.2 <i>Diametro Principale</i>	12
3.1.3 <i>Diametro Secondario</i>	12
3.1.4 <i>Numero di foglie</i>	13
3.1.5 <i>Lunghezza delle foglie</i>	13
3.1.6 <i>Numero di fiori</i>	14
<b>COMMENTO</b>	14

3.2 Le variabili nelle aree del campo nelle differenti condizioni di luce.....	15
3.2.1 Altezza .....	16
3.2.2 Diametro Principale .....	17
3.2.3 Diametro Secondario .....	18
3.2.4 Numero di foglie .....	19
3.2.5 Lunghezza delle foglie .....	20
3.2.6 Numero di fiori .....	21
3.3 Condizioni di luce confrontate tra loro (Analisi Bivariata) .....	22
3.3.1 La correlazione delle variabili nelle differenti condizioni di luce.....	22
3.3.2 Altezza ( <i>height</i> ) influenzata dal numero di foglie ( <i>number_leaves</i> ) .....	23

<b>4. MODELLISTICA</b> .....	<b>27</b>
4.1 Modelli lineari generalizzati (GLM) .....	27
4.2 Modelli di regressione polinomiale.....	28
4.3 Modelli additivi generalizzati (GAM) .....	29
4.4 Applicazione dei modelli per ciascuna variabile .....	32
4.4.1 Altezza .....	33
4.4.2 Diametro Principale .....	34
4.4.3 Diametro Secondario .....	35
4.4.4 Numero di foglie .....	36
4.4.5 Lunghezza delle foglie .....	37
4.4.6 Numero di fiori.....	38
<b>COMMENTO</b> .....	<b>39</b>

<b>5. DIFFERENZA TRA LE CONDIZIONI DI LUCE</b> .....	41
5.1 ANOVA a un fattore .....	42
5.2 ANOVA a due fattori .....	43
5.3 ANCOVA .....	44
5.4 Test ANCOVA per ciascuna variabile .....	47
5.4.1 Altezza .....	48
5.4.2 Diametro Principale .....	49
5.4.3 Diametro Secondario .....	50
5.4.4 Numero di foglie .....	51
5.4.5 Lunghezza delle foglie .....	52
5.4.6 Numero di fiori .....	53
<b>COMMENTO</b> .....	54
<b>6. I FATTORI METEO</b> .....	55
6.1 Analisi delle componenti principali (PCA) .....	55
6.2 Variabili del dataset Meteo .....	61
6.3 Applicazione della PCA per ciascuna condizione di luce .....	62
6.3.1 Altezza .....	63
6.3.2 Diametro Principale .....	64
6.3.3 Diametro Secondario .....	65
6.3.4 Numero di foglie .....	66
6.3.5 Lunghezza delle foglie .....	68
6.3.6 Numero di fiori .....	69
<b>COMMENTO</b> .....	71
<b>CONCLUSIONI</b> .....	73
<b>BIBLIOGRAFICA</b> .....	75





# 1. INTRODUZIONE

## 1.1 L'impianto fotovoltaico

Il termine fotovoltaico è composto dal termine greco *photo*, ovvero luce, e dall'aggettivo *voltaico*, comunemente utilizzato per indicare l'energia elettrica, coniato in onore di Alessandro Volta. L'impianto fotovoltaico è attualmente il mezzo più diffuso per generare energia sostenibile sfruttando il Sole, il quale permane la fonte energetica rinnovabile più disponibile e abbondante sul nostro Pianeta.

L'idea di fondo è di esaurire i combustibili fossili (petrolio, carbone e gas) per prediligere esclusivamente il Sole, il vento, le maree ecc, cioè fonti sostenibili in grado di rigenerarsi in tempi brevi, quindi tecnicamente non esauribili, e che permettono di ridurre le emissioni di CO<sub>2</sub> nell'atmosfera.

### 1.1.1 Invenzione del fotovoltaico

L'impianto sfrutta l'effetto fotovoltaico, scoperto nel 1839 dal giovane fisico francese Alexandre Edmond Becquerel durante un esperimento che prevedeva di immergere due elettrodi metallici in una soluzione conduttrice. Becquerel notò che i due conduttori generavano un flusso di corrente più intenso se esposti alla luce, dunque concluse che una piccola quantità di energia corrente viene prodotta durante alcune reazioni chimiche indotte dalla luce solare.

Qualche anno più tardi, nel 1876, gli scienziati Richard Evans, Willoughby Smith e William G. Adams furono capaci di dimostrare che il selenio produce una lieve scarica di energia se esposto alla luce; la scoperta ispirò l'americano Charles Fritts e nel 1879 creò il primo pannello fotovoltaico composto da uno strato di selenio avvolto in una pellicola sottile d'oro. Successivi studi e in seguito ad accurate ricerche, nel 1954, il selenio venne sostituito dal silicio, un semimetallo più economico e più efficace; la produzione di pannelli si intensificò notevole attribuendo in questo modo ai pannelli il titolo di principali mezzo per la conversione di energia solare in energia elettrica<sup>1</sup>.

<sup>1</sup> Coiante, D. (2008). *Fotovoltaico. Il processo evolutivo e le nuove frontiere*. ENEA

### 1.1.2 Composizione e funzionamento dell'impianto fotovoltaico

Un impianto fotovoltaico si compone sostanzialmente di un sistema che permette la conversione della potenza denominato *inverter*, di un accumulatore di energia, ossia la batteria (*battery bank*), e di pannelli fotovoltaici che fungono da generatori di energia (*solar panels*). I singoli pannelli sono un insieme di celle ognuna delle quali consente la conversione dell'energia solare catturata in energia elettrica; essi sono il componente elementare dell'impianto, presentano una forma rotonda oppure quadrata, una superficie di circa 0.3 mm e sono realizzate con materiale semiconduttore, spesso il silicio.

L'energia solare è costituita da fotoni, quando uno di questi colpisce la superficie di una cella dell'impianto si attua l'effetto fotovoltaico, ossia il processo che genera energia grazie ai fotoni; in pratica quando un fotone raggiunge la cella, l'energia solare al suo interno viene rilasciata e trasferita agli elettroni presenti nella cella in silicio, i quali vengono attivati e cominciano a fluire nel circuito producendo così energia elettrica. Il processo permette la generazione di energia sottoforma di corrente continua, in seguito l'*inverter* avrà il compito di modificarla in corrente alternata, la quale viene poi trasferita e utilizzata nelle reti di distribuzione.

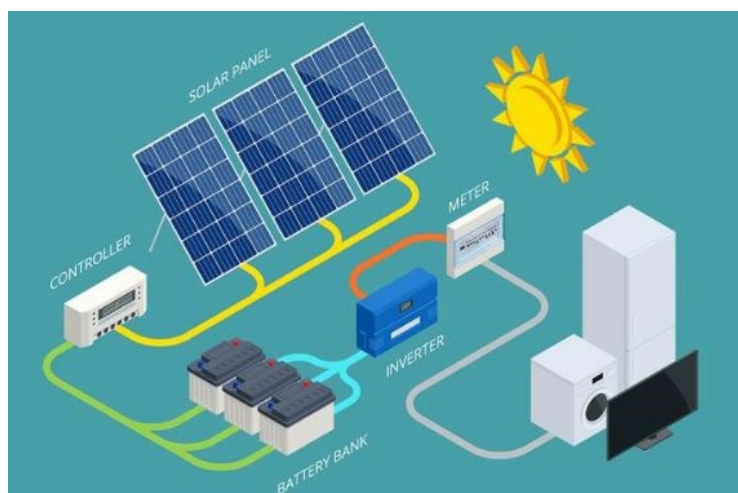


Figura 1a: Impianto fotovoltaico scomposto nelle sue componenti<sup>2</sup>

<sup>2</sup>Fonte: *Da quali parti è formato un impianto fotovoltaico? Alla scoperta delle componenti principali!* (2016, August 21). Solandria. <https://www.solandria.it>

### 1.1.3 Sustainable Development Goals

A New York, in occasione della settimana del clima è stato azionato un enorme orologio digitale sulla parete di un edificio a Union Square; l'orologio, chiamato "Climate Clock", segna due numeri: il countdown al superamento del limite soglia del surriscaldamento globale fissata al 1.5 °C se il mondo continuasse a produrre l'eccessiva emissione nell'atmosfera di CO<sub>2</sub> attuale, ossia il tempo che rimane per intraprendere azioni incisive per mantenere il surriscaldamento globale sotto la soglia e la percentuale di energia generata da fonti sostenibili in tutto il mondo, una sorta di speranza e di monitoraggio delle soluzioni attualmente più adeguate per allungare il countdown.

In concomitanza con la realizzazione del "Climate Clock", il 25 settembre 2015, i 193 Paesi membri delle Nazioni Unite hanno sottoscritto un programma che vede interessati 17 obiettivi per lo Sviluppo Sostenibile (Sustainable Development Goals); programma successivamente approvato dall'assemblea Generale dell'ONU e denominato "Agenda 2030", in quanto si spera di raggiungerli entro l'anno 2030. L'incremento di energia rinnovabile, l'aumento di infrastrutture sostenibili e l'accesso a servizi energetici a prezzi accessibili è trattato nell'obiettivo numero 7 dell'Agenda 2030.

A ogni Paese membro viene chiesto di sviluppare la propria strategia d'azione, in Italia tale strategia viene gestita dalla Cabina di regia denominata "Benessere Italia", organo della Presidenza del Consiglio a cui spetta il compito di monitorare e in caso migliorare le politiche volte al raggiungimento degli obiettivi prefissati<sup>3</sup>.



Figura 2a: Climate clock. Domenica, 12 febbraio 2023 (CET) alle ore 15:00 (PADOVA)<sup>4</sup>

<sup>3</sup>Turano, V. (2023, January 16). *Agenda 2030 per lo sviluppo sostenibile*. Agenzia per La Coesione Territoriali. <https://www.agenziacoesione.gov.it/comunicazione/agenda-2030-per-lo-sviluppo-sostenibile/>

<sup>4</sup>Fonte: *Get Your Clock*. (a.n.d.-b). Climate Clock. <https://climateclock.world/clocks>

## 1.2 Oggetto di studio

L'oggetto di studio trattato è una simulazione che combina dei pannelli fotovoltaici alla coltivazione di culture a terra, più precisamente alla coltivazione delle patate. Si definisce così un nuovo fenomeno che prende il nome di impianto agrivoltaico, che a differenza del tradizionale fotovoltaico, installato a lato dei campi coltivati, è pensato per limitare e sfruttare al meglio il terreno, infatti è un sistema sospeso sopra i campi che poggia su due trampoli sui quali è installato l'asse principale e gli assi secondari che a loro volta sostengono i pannelli.



Figura 3a: Impianto fotovoltaico tradizionale a sx e impianto agrivoltaico a dx

Il vantaggio dell'agrivoltaico è l'aumento dei campi disponibili alla coltivazione, oltre al fatto che protegge il prodotto dagli eventi meteorologici estremi. Inoltre l'impianto inumidisce il terreno generando delle zone d'ombra, di conseguenza la pianta è in grado di assorbire la quantità necessaria alla traspirazione limitando così il consumo idrico.

Il concetto di agrivoltaico venne coniato per la prima volta nel 1981 dal tedesco Adolf Goetzberger, all'epoca direttore di un'organizzazione per lo sviluppo della ricerca applicata; Goetzberger espose la sua idea dandone un assetto teorico dimostrandone l'efficacia e cercando di incentivare le aziende agricole al suo utilizzo, ma i costi per la creazione degli impianti fotovoltaici era ancora troppo alto. I primi prototipi si iniziarono a realizzare solamente negli anni 2000, prima in Giappone, in cui l'ingegnere Akira Nagashima iniziò a sviluppare prototipi per scopi sperimentali, successivamente in Europa, soprattutto in Italia, nelle province di Mantova, Piacenza e Verona, in cui gli impianti assunsero un ruolo fondamentale per le aziende agricole<sup>5</sup>.

<sup>5</sup>Amaducci, S., Yin, X., & Colauzzi, M. (2018). Agrivoltaic systems to optimise land use for electric energy production. *Applied Energy*, 220, 545-561. <https://doi.org/10.1016/j.apenergy.2018.03.081>

### **1.3 Raccolta dei dati**

La ricerca si è svolta nella località di Piacenza, in Emilia-Romagna, per una durata complessiva di dieci settimane, dal 28 aprile 2021 al 05 luglio 2021. Una prima nota informativa riguarda il periodo di osservazione perché le osservazioni non sono state rilevate ogni settimana precisa, ma con una unità di misura indipendente dalle settimane come le si conoscono; il giorno scelto per le rilevazioni non coincide per ciascuna settimana e le settimane non sono consecutive tra loro, ma tra tutte intercorre indicativamente il medesimo lasso di tempo; si può presupporre quindi che differenze significative utili per le analisi si ottengono solamente dopo un determinato lasso di tempo.

Il campo soggetto a rilevazione è suddiviso in quattro aree della stessa dimensione e da ognuna viene estratta una piantina con un campionamento casuale semplice; tale metodo prevede un'estrazione del tutto casuale dal momento che ogni piantina ha la stessa probabilità di essere selezionata. Le singole piantine dunque sono le unità statistiche osservate e sono selezionate all'interno del campo con un campionamento stratificato; tale tecnica garantisce una rilevazione consona con l'analisi che si vuole condurre, in quanto cerca di captare le piantine in ogni zona significativa per l'analisi, garantendo così la formazione di un campione rappresentativo. Per semplicità si decide di assumere che ogni settimana venga estratta la stessa piantina, anche se in realtà ogni settimana viene estratta una piantina differente rispetto alla settimana precedente. La raccolta dei dati avviene direttamente sul campo permettendo così di ottenere dati diretti, dunque non generati da stime.

Le rilevazioni vengono effettuate in quattro diversi campi, uno privo di alcun impianto (FL), ossia soggetto a luce solare piena, e tre che differiscono tra loro in base alle caratteristiche dell'impianto agrivoltaico installati sopra. Gli impianti si distinguono tra loro principalmente per la quantità di assi secondari, montati perpendicolarmente ai principali, e quindi per la quantità di ombra che generano sul campo. Una seconda caratteristica rilevante è che i pannelli non sono distribuiti in modo uniforme sulla superficie del terreno, dunque generano una zona d'ombra lungo l'asse principale più intensa e una seconda che si verifica solamente in alcuni istanti della giornata<sup>6</sup>.

L'impianto agrivoltaico denominato SH30 simula la versione 1 del modello con pannelli stabili, costituiti da 5 assi secondari e 10 pannelli solari che producono un ombreggiamento medio costante del terreno sottostante del 30%.

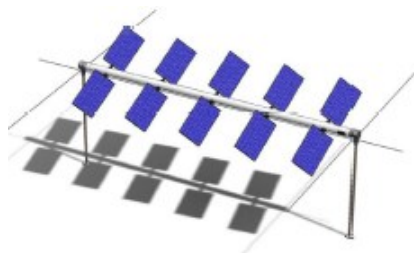


Figura 4a: Impianto agrivoltaico SH30<sup>6</sup>

L'impianto SH60 invece simula la versione 2, che a differenza della versione 1, è in grado di seguire la luce solare mediante dei motori elettrici interconnessi; essi presentano 4 assi secondari e 32 pannelli solari, capaci di generare una zona d'ombra costante del 60%.

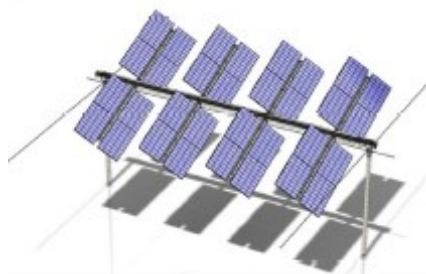


Figura 5a: Impianto agrivoltaico SH60<sup>6</sup>

Infine l'impianto HSH60 non è fisicamente realizzabile perché è una condizione di luce che simula la versione 2 parzialmente, ossia è la zona che si torva nel mezzo di due track dell'impianto SH60<sup>6</sup>.

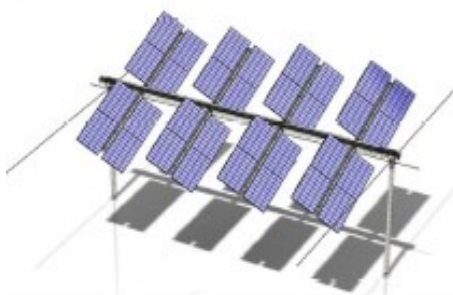


Figura 6a: Condizione di luce HSH60<sup>6</sup>

<sup>6</sup>Amaducci, S., Yin, X., & Colauzzi, M. (2018). Agrivoltaic systems to optimise land use for electric energy production. *Applied Energy*, 220, 545-561. <https://doi.org/10.1016/j.apenergy.2018.03.081>

#### **1.4 Obiettivi dell'analisi**

La stesura della tesi si divide principalmente in due categorie; si procede con un'analisi esplorativa, univariata e bivariata, per cogliere le differenze significative delle caratteristiche (variabili) delle unità statistiche, ossia delle piantine, nelle diverse condizioni di luce.

Successivamente si intraprende la tematica della modellazione delle curve di crescita delle unità statistiche nei diversi campi. Un primo obiettivo consiste nel cogliere la relazione nel tempo delle variabili interessate, servendosi dei modelli GAM. Un secondo obiettivo prevede invece di determinare per ciascuna variabile se esistono differenze significative tra le varie condizioni di luce mediante il test *ANOVA*. Infine per il terzo obiettivo si introduce un secondo dataset contenenti le variabili che riguardano il meteo per stabilire se le condizioni meteorologiche influenzano in qualche modo la crescita delle unità; l'obiettivo consiste nel trovare la posizione dell'impianto agrivoltaico che ottimizza la crescita, ossia più vicina alla condizione di luce piena (FL), considerata la situazione ottimale. Per l'ultimo obiettivo si utilizza l'analisi delle componenti principale (PCA) perché le variabili meteo risultanti sono tante in confronto alle osservazioni.





## 2. IL DATASET E DELLE VARIABILI

### 2.1 Potatoes

Il dataset in questione riporta le osservazioni in un giorno arbitrario della settimana; il giorno viene rappresentato dalla variabile “**date**” mentre la settimana dalla variabile “**week**”. Per ogni settimana selezionata vengono considerati quattro campi indipendenti tra di loro che rappresentano la variabile “**light**”, ciascuno soggetto a una condizione di luce differente da quelle degli altri campi.

- **FL** condizione di luce piena
- **SH30** condizione sotto l’agrivoltaico con ombreggiamento medio costante del 30%
- **SH60** condizione dell’agrivoltaico con ombreggiamento medio costante del 60%
- **HSH60** condizione SH60 ma parziale, ossia nel mezzo di due track

Per ogni condizione di luce si suddivide il campo in questione in quattro aree della stessa dimensione; tale suddivisione verrà contrassegnata dalla variabile “**area**”. Successivamente per le analisi si assumerà che nel corso delle rilevazioni le piantine siano le medesime indicate della variabile “**plant**”.

Ogni piantina rappresenta un’osservazione e in quanto tale vengono riportate le sue caratteristiche:

- **height**: altezza della piantina fuori dalla terra in centimetri
- **diameter1**: lunghezza dell’asse principale rivolto verso nord (piantina vista dall’altro)
- **diameter2**: lunghezza dell’asse secondario, perpendicolare al principale
- **number\_leaves**: numero di foglie
- **length\_leaves**: lunghezza delle foglie
- **number\_flowers**: numero dei fiori

## 2.2 Meteo

Il secondo dataset “Meteo” viene utilizzato nella sezione che vede come oggetto di studio la modellazione delle curve di crescita delle piantine; i dati che lo compongono sono stati registrati da alcune centraline installate in prossimità dei campi. Si hanno a disposizione quattro dataset distinti, suddivisi in base alla condizione di luce a cui è soggetta il campo (FL, SH30, SH60 e HSH60), tutti riportano misurazioni temporali delle seguenti caratteristiche:

- **temp\_aria**: temperatura dell’aria
- **umid\_aria**: umidità dell’aria
- **p\_rugiada**: punto di rugiada
- **umid\_suolo**: umidità del suolo
- **temp\_suolo**: temperatura del suolo
- **rad\_solare2**: radiazione solare

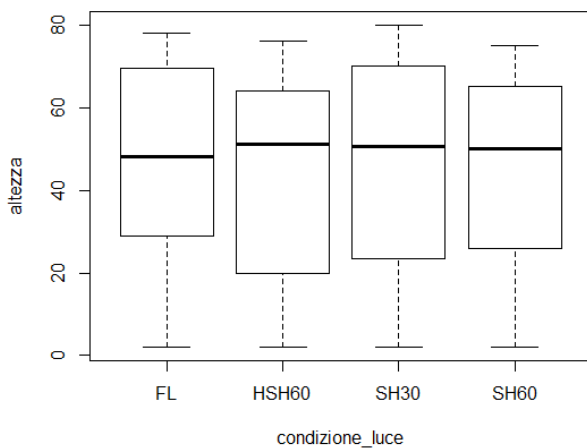
La variabile “**data**” contiene il giorno e l’ora della rilevazione; vengono effettuate osservazioni ogni ora dalle 00.00 alle 23:00 di ogni giorno dal 21 aprile 2021 al 05 luglio 2021. Per la temperatura dell’aria è stata calcolata la sua media oraria, quindi inserita nella variabile “**temp\_aria\_media**” e il suo valore massimo e minimo registrato nell’ora, identificati rispettivamente dalle variabili “**temp\_aria\_max**” e “**temp\_aria\_min**”. Le medesime rilevazioni sono state svolte per l’umidità dell’aria (“**umid\_aria**”), per il punto di rugiada (“**p\_rugiada**”), per l’umidità e per la temperatura del suolo (“**umid\_suolo**” e “**temp\_suolo**”). Il punto di rugiada indica la temperatura in cui l’aria diventa satura di vapore acqueo cominciando così a trasformarlo in gocce d’acqua. Vengono inoltre introdotte le variabili che riguardano la radiazione solare, ossia l’energia radiante emessa dal Sole, viene così definita la media delle radiazioni solari nella variabile “**rad\_solare2\_media**”, il totale nella variabile “**rad\_solare2\_tot**” e le ore nella variabile “**rad\_solare2\_ore**”; infine il valore massimo e minimo delle radiazioni sono contenuti rispettivamente nelle variabili “**rad\_solare2\_max**” e “**rad\_solare2\_min**”. L’unità di misura della radiazione solare è  $\text{kJ/m}^2$ .

## 3. ANALISI ESPLORATIVE

### 3.1 Condizioni di luce confrontate tra loro (Analisi Univariata)

Una prima analisi sulle piantine è volta a scrutare la differenza delle singole variabili nelle varie condizioni di luce, mediante l'ausilio di un boxplot per ciascuna variabile interessata. Si procede quindi con un'analisi univariata, ossia considerando una variabile per volta. Un primo accorgimento ricade sul numero di osservazioni che risulta essere in totale 384; dunque nello specifico ciascuna condizione di luce dispone di 96 rilevazioni, infatti vengono selezionate 16 piantine per ogni istante temporale, per un periodo di osservazione complessivo di osservazione pari a sei settimane.

#### 3.1.1 Altezza

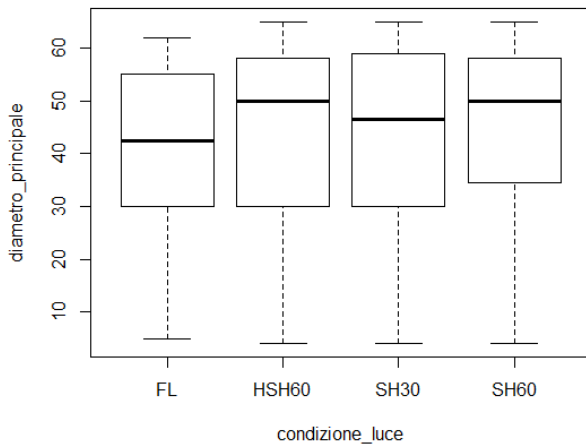


	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
<b>FL</b>	2.00	29.00	48.00	45.36	69.25	78.00
<b>HSH60</b>	2.00	20.00	51.00	42.65	63.50	76.00
<b>SH30</b>	2.00	23.75	50.50	45.45	70.00	80.00
<b>SH60</b>	2.00	26.00	50.00	44.30	65.00	75.00

Figura 7a: Boxplot della variabile height

Nessuno dei boxplot contiene outlier, ossia valori anomali. La condizione SH30 sembrerebbe consentire una crescita media delle piantine leggermente superiore (45.45) rispetto a quella delle piantine situate nei campi non citati, rispettivamente con una crescita media pari a 45.36, 42.65 e 44.30. Le piantine del campo SH30 inoltre assumono una crescita massima (80.00) maggiore rispetto alle altre piantine, quindi la crescita è favorita dalla presenza di un impianto agrivoltaico che genera una quantità d'ombra limitata, in questo caso pari al 30%.

### 3.1.2 Diametro Principale

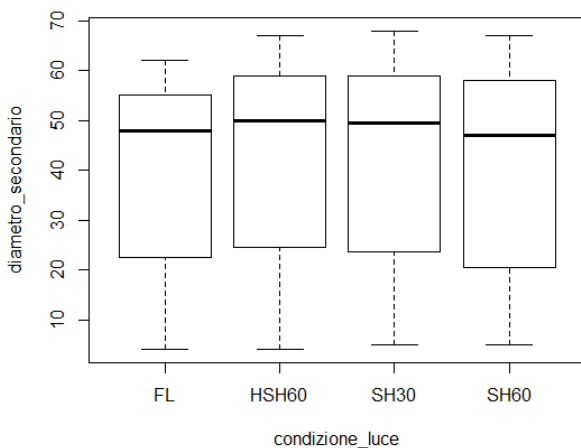


	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
<b>FL</b>	5.00	30.00	42.50	39.60	55.00	62.00
<b>HSH60</b>	4.00	30.00	50.00	41.33	58.00	65.00
<b>SH30</b>	4.00	30.75	46.50	41.72	59.00	65.00
<b>SH60</b>	4.00	34.75	50.00	42.39	58.00	65.00

Figura 8a: Boxplot della variabile diameter1

Le lunghezze di dei segmenti superiori del boxplot sono differenti delle lunghezze dei segmenti inferiori, si può affermare dunque che il 25% dei valori inferiori si disperde in misura maggiore della corrispondente percentuale dei valori più elevati indipendentemente dalla condizione di luce visionata; dispersione che risulta essere leggermente maggiore nella condizione SH60.

### 3.1.3 Diametro Secondario

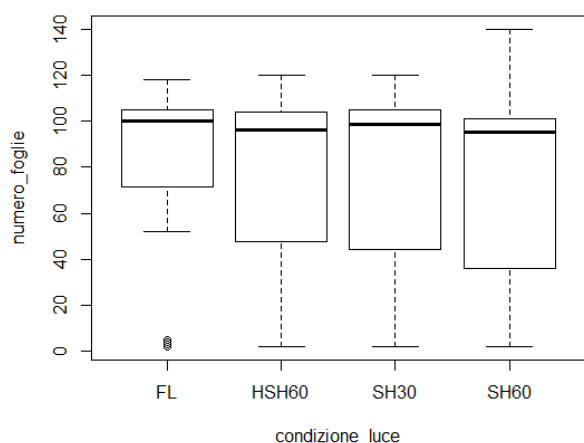


	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
<b>FL</b>	4.00	22.75	48.00	39.35	55.00	62.00
<b>HSH60</b>	4.00	24.75	50.00	41.03	59.00	67.00
<b>SH30</b>	5.00	23.75	49.50	41.98	59.00	68.00
<b>SH60</b>	5.00	20.75	47.00	39.51	58.00	67.00

Figura 9a: Boxplot della variabile "diameter2"

I valori centrali delle distribuzioni mostrano una notevole variabilità e si distribuiscono attorno a una mediana spostata verso l'alto, ossia verso valori elevati. Tale variabilità può essere catturata mediante il range interquartile che risulta essere pari rispettivamente a 32.25, 35.25, 37.25 e 34.25; anche il diametro secondario presenta una variabilità maggiore nella condizione SH60.

### 3.1.4 Numero di foglie

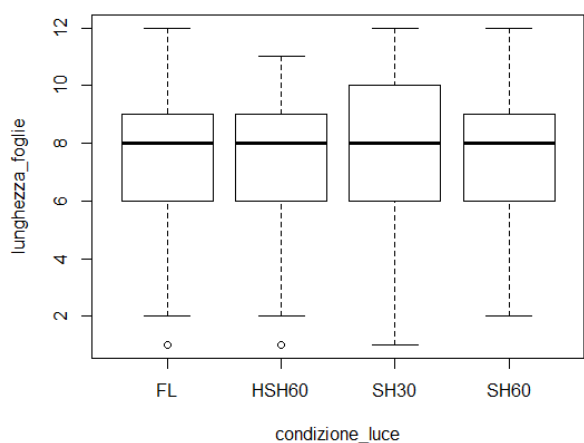


	Min	1st Qu	Median	Mean	3rd Qu.	Max
FL	2.00	71.75	100.00	80.97	105.00	118.00
<b>HSH60</b>	2.00	49.75	96.00	74.88	104.00	120.00
<b>SH30</b>	2.00	44.75	98.50	75.97	105.00	120.00
<b>SH60</b>	2.00	36.00	95.00	74.03	101.00	140.00

Figura 10a: Boxplot della variabile "number\_leaves"

La condizione FL contiene dei valori anomali, nettamente inferiori, che possono essere causati da eventi meteorologici estremi. FL e SH30 permettono alle piantine di sviluppare mediamente più foglie, infatti il 50% delle osservazioni risultano avere rispettivamente più di 100.00 e di 98.50 foglie; dunque una maggiore esposizione alla luce permette un maggior sviluppo di foglie.

### 3.1.5 Lunghezza delle foglie

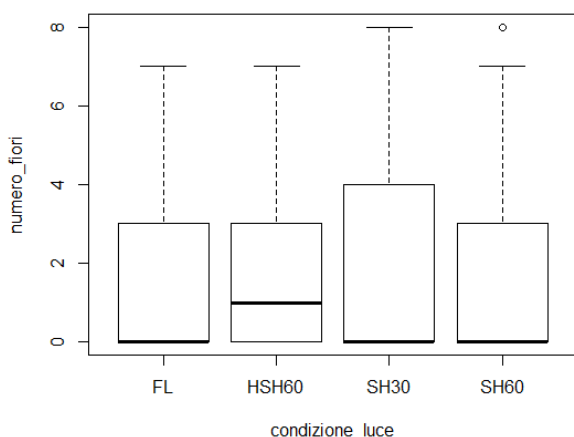


	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
<b>FL</b>	1.00	6.00	8.00	7.07	9.00	12.00
<b>HSH60</b>	1.00	6.00	8.00	6.97	9.00	11.00
<b>SH30</b>	1.00	6.00	8.00	7.69	10.00	12.00
<b>SH60</b>	1.00	6.00	8.00	7.21	9.00	12.00

Figura 11a: Boxplot della variabile "length\_leaves"

La lunghezza delle foglie nelle differenti condizioni di luce è una distribuzione che poggia attorno alla medesima mediana, localizzata in prossimità del valore 8.00. Le varie distribuzioni si distribuiscono in modo molto simile tra loro, anche se i casi FL e SH30, tendono ad assumere una lunghezza maggiore; quindi la luce solare favorisce anche la lunghezza delle foglie.

### 3.1.6 Numero di fiori



	Min.	1st Qu.	Media n	Mea n	3rd Qu.	Max
<b>FL</b>	0.00	0.00	0.00	1.92	3.00	7.00
<b>HSH60</b>	0.00	0.00	1.00	1.69	3.00	7.00
<b>SH30</b>	0.00	0.00	0.00	2.03	4.00	8.00
<b>SH60</b>	0.00	0.00	0.00	1.60	3.00	8.00

Figura 12a: Boxplot della variabile "number\_flowers"

I boxplot riguardanti il numero di fiori riportano una forte simmetria a destra, ossia positiva; suggerisce che mediamente le piantine sviluppano pochi fiori; inoltre osservando i vari diagrammi si intuisce che la condizione SH30 è la migliore per lo sviluppo, in quanto il terzo quartile (4.00) risulta maggiore dei corrispettivi quantili nelle distribuzioni nelle altre condizioni di luce (3.00).

## COMMENTO

In generale si sono ottenuti risultati migliori, dunque valori maggiori nelle condizioni influenzate da un impianto agrivoltaico, anche se SH30 rimane comunque la condizione migliore rispetto alle altre condizioni influenzate da un impianto. La combinazione dei fotovoltaici con la coltivazione migliora le performance sotto molti aspetti, quali la crescita, lo sviluppo della lunghezza delle foglie e di un numero maggiore di fiori. L'incremento del numero delle foglie, invece, rimane migliore nella condizione di luce base (FL), dunque le foglie necessitano di una quantità di luce considerevole per svilupparsi e per non cadere.

### 3.2 La crescita delle piantine nelle differenti condizioni di luce

Una seconda analisi descrittiva è volta a monitorare le caratteristiche della piantina nell'arco temporale, ossia nelle sei settimane di osservazione, considerando le condizioni di luce del campo.

Il campo sottoposto a rilevazione è suddiviso in quattro aree delle medesime dimensioni e da ciascuna area vengono estratte quattro piantine; si riporta in seguito la rappresentazione di uno dei campi, quello sottoposto a luce piena (FL), tale partizione viene svolta anche per gli altri campi.

Figura 13a: Rappresentazione del campo in condizione di luce piena (FL)

11FL	21FL	12FL	22FL
31FL	41FL	32FL	42FL
13FL	23FL	14FL	24FL
33FL	43FL	34FL	44FL

Di fatto vengono selezionate 16 piantine per ciascun campo, quindi in sei settimane avremo in totale 96 osservazioni per ogni condizione di luce.

Per l'analisi si considera la media delle 4 piantine contenute in un'area del campo, avendo assunto precedentemente che presentano caratteristiche simili se si trovano all'intero della stessa area; l'unità di misura ora diventa l'area del campo. Si ottengono così quattro dataset suddivisi in base alla condizione di luce e ciascuno dispone di 24 osservazioni.

#### PRIMA SETTIMANA

$$\text{PRIMA AREA (FL)} \rightarrow 1\text{FL}_{\text{prima}} = \frac{1}{4} \sum_{i=1}^4 FL_{i1}$$

$$\text{SECONDA AREA (FL)} \rightarrow 2\text{FL}_{\text{seconda}} = \frac{1}{4} \sum_{i=1}^4 FL_{i2}$$

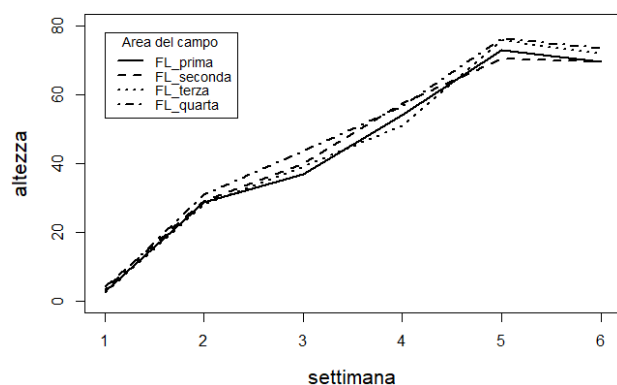
$$\text{TERZA AREA (FL)} \rightarrow 3\text{FL}_{\text{terza}} = \frac{1}{4} \sum_{i=1}^4 FL_{i3}$$

$$\text{QUARTA AREA (FL)} \rightarrow 4\text{FL}_{\text{quarta}} = \frac{1}{4} \sum_{i=1}^4 FL_{i4}$$

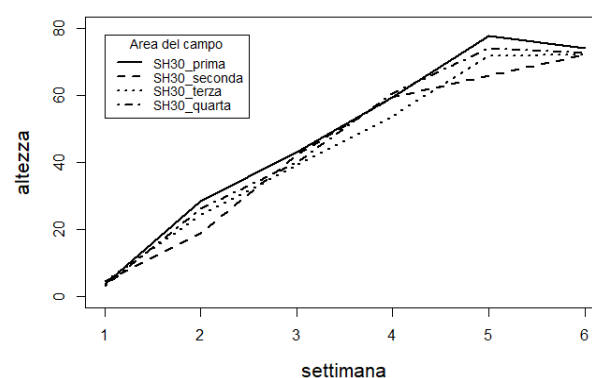
Si riportano i grafici delle variabili in ogni area del campo nelle condizioni di luce.

### 3.2.1 Altezza

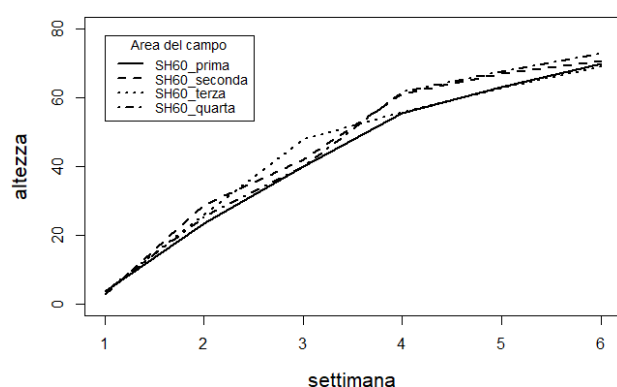
**FL**



**SH30**



**SH60**



**HSH60**

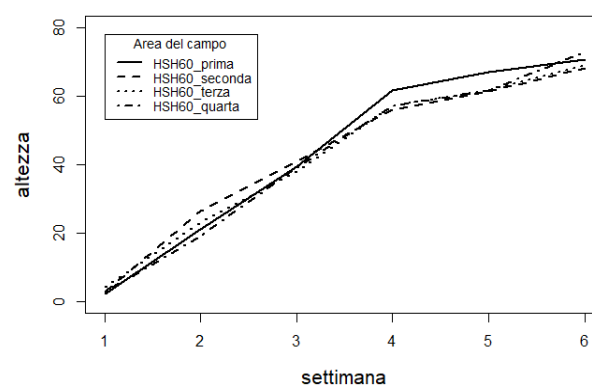


Figura 14a: Curve dell'altezza (height) per ogni area nelle differenti condizioni di luce

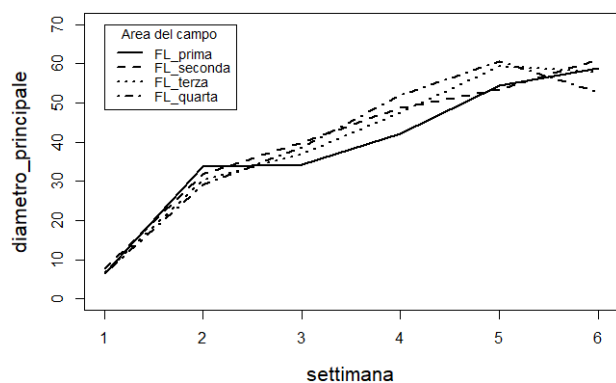
Le condizioni di luce sono favorevoli per la crescita delle piante che indipendentemente dalla condizione a cui il campo è sottoposto, varia in ogni area, soprattutto nelle condizioni influenzate dall'impianto. Questo dipende dalla posizione dell'agrivoltaico che altera l'ombreggiatura e di conseguenza l'umidità del terreno in alcune aree.

A primo impatto la condizione di luce piena (FL) e la condizione SH30 sembrano avere un andamento di crescita simile rispetto alle condizioni SH60 e HSH60. Anche se la seconda area di SH30 sembra affiancarsi più a quest'ultime, in quanto influenza dall'ombra dell'agrivoltaico. L'andamento delle prime due condizioni raggiunge un punto di svolta attorno alla quinta settimana, invece le altre due attorno alla quarta settimana; dunque una quantità di luce solare permette una crescita maggiore per le piantine. Tale svolta determina un cambiamento notevole in termini di pendenza e marca una profonda differenza tra le prime due condizioni e ultime due; le prime, raggiunto tale punto, iniziano una lieve decrescita, le altre invece continuano la crescita anche se con una pendenza inferiore.

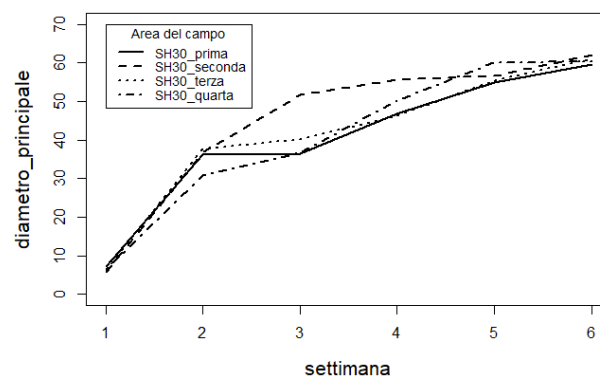


### 3.2.2 Diametro Principale

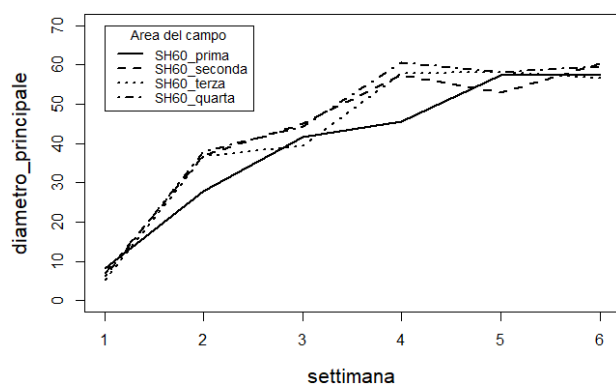
**FL**



**SH30**



**SH60**



**HSH60**

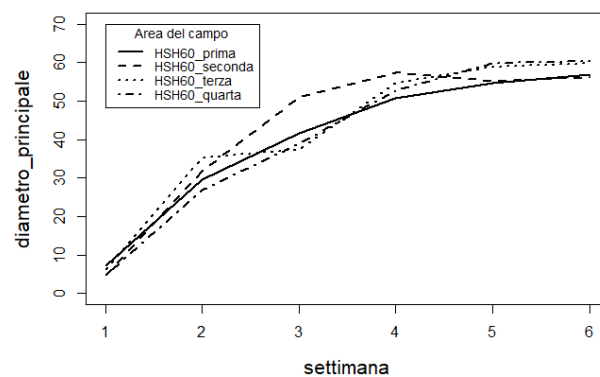


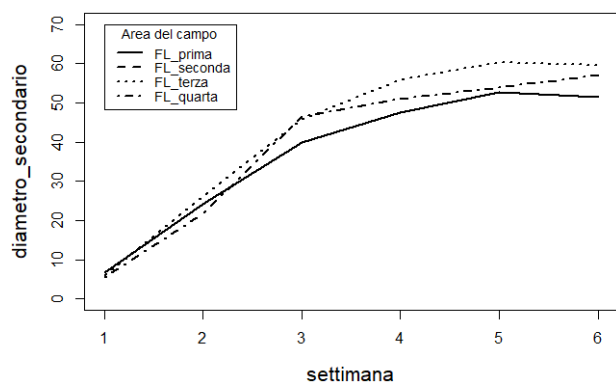
Figura 15a: Curve del diametro principale (*diameter1*) per ogni area nelle differenti condizioni di luce

La crescita vertiginosa nella fase iniziale delle piantine in FL è simile all'andamento nei campi SH30 e HSH60, inoltre la prima area dei campi FL e SH60 si discosta dall'andamento generale per poi avvicinarsi alle altre curve attorno alla sesta settimana. Nel complesso le aree del campo FL riportano un andamento di crescita costante fino all'ultima settimana osservata e il motivo dell'assomiglianza è che il campo è colpito dalla stessa quantità di luce in tutti i suoi punti; in SH60 invece vi è un'elevata quantità di ombra che si distribuisce equamente infatti le curve crescono in modo non regolare ma in modo simile tra loro a eccezione della prima curva che mostra un andamento di crescita distaccato.

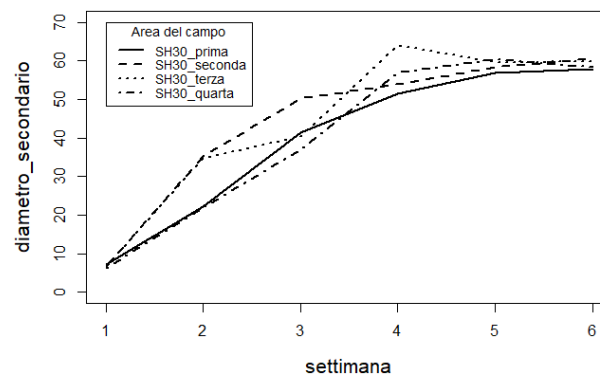
In SH30 e in HSH60 invece la seconda area di entrambe le condizioni sembra discostarsi dall'andamento generale, soprattutto nelle fasi intermedie dello sviluppo, in cui assume valori elevati rispetto a quelli assunti nelle altre aree. Tali campi sono soggetti a una minor quantità d'ombra che colpisce quasi esclusivamente quest'area, permettono così alle piantine contenute al suo interno di sviluppare un dinametro maggiore.

### 3.2.3 Diametro Secondario

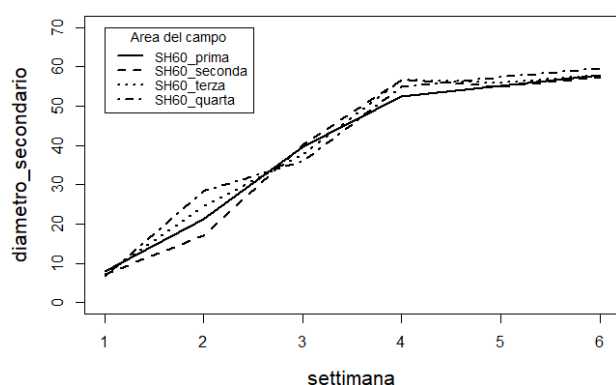
FL



SH30



SH60



HSH60

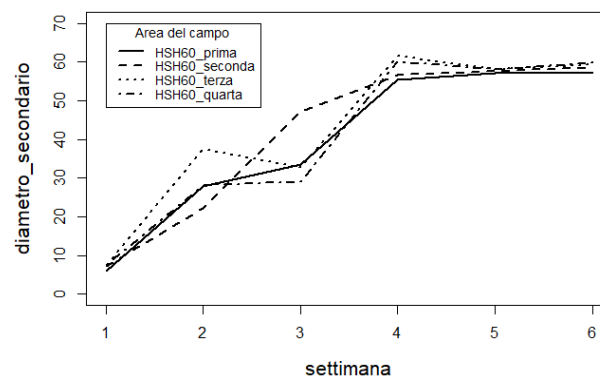


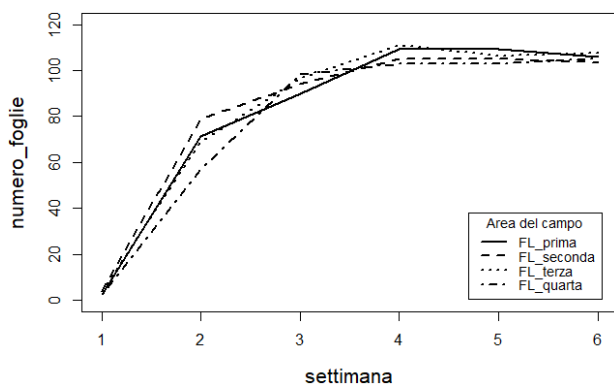
Figura 16a: Curve del diametro secondario (*diameter2*) per ogni area nelle differenti condizioni di luce

Anche le curve del diametro secondario nelle diverse aree del campo sono coerenti nella condizione FL e SH60, anche se nella prima settimana la prima e la terza area del campo SH60 si differenziano dalla seconda e dalla quarta. SH60 genera un ambiente tale da permettere alle piantine di sviluppare maggiormente il loro diametro entro le 4 settimane, invece in FL la curva di crescita varia il suo andamento una settimana prima, attorno alla terza settimana. In SH30 nelle seconda e nella terza area le piantine sviluppano un diametro nettamente superiore a quello delle piantine nella prima e nella quarta area, nella fase iniziale; in HSH60 invece le piantine nella terza area assumono sempre valori maggiori e sembrano seguire il medesimo andamento della prima e della quarta area; la seconda area invece assume un andamento differente.

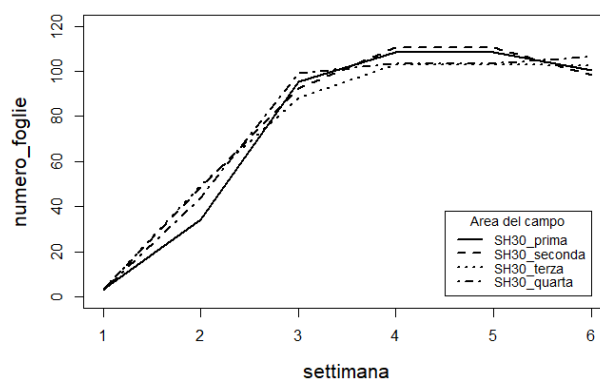
In generale i campi soggetti a un impianto agrivoltaico contengono piantine che continuano a sviluppare il loro diametro secondario anche dopo la sesta settimana; invece in FL tale crescita si arresta attorno alla quinta settimana per poi decrescere.

### 3.2.4 Numero di foglie

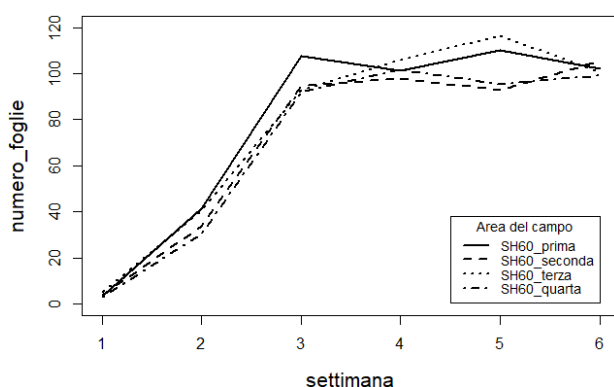
**FL**



**SH30**



**SH60**



**HSH60**

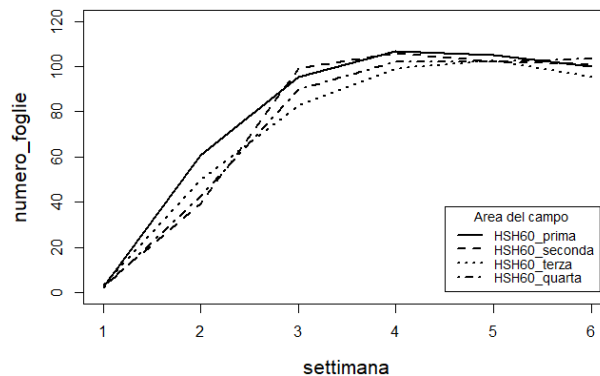


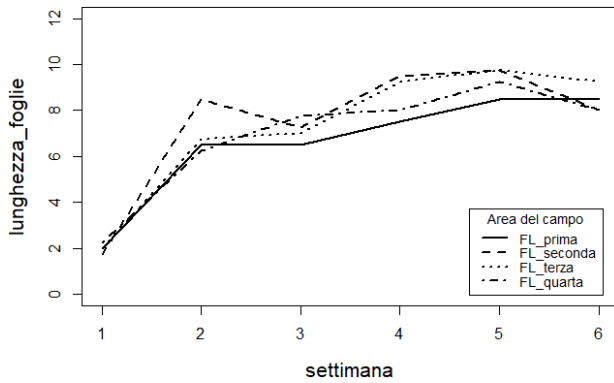
Figura 17a: Curve del numero di foglie (*number\_leaves*) per ogni area nelle differenti condizioni di luce

Le piantine disposte in campi soggetti a luce piena (FL) generano una quantità maggiore di foglie rispetto quella generata dalle piantine influenzate da un impianto agrivoltaico, infatti le foglie necessitano di una notevole dose di luce solare per svilupparsi. L'andamento del numero di foglie subisce una lieve stabilizzazione attorno alla seconda settimana in FL e attorno alla terza settimana nelle condizioni di luce con l'agrivoltaico.

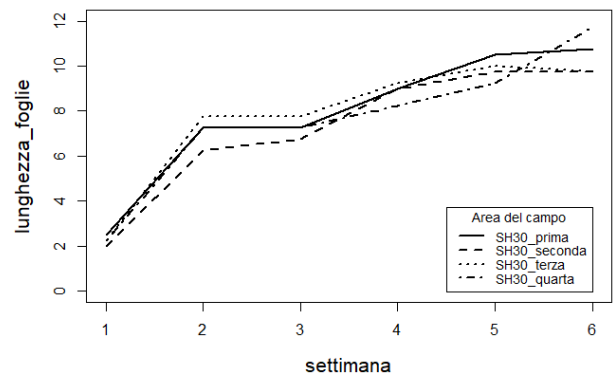
La crescita rapida del numero di foglie si verifica in FL ma anche in SH30, infatti anche questa condizione è soggetta da una quantità di luce non indifferente. In SH60 invece la crescita avviene più lentamente soprattutto nella prima settimana.

### 3.2.5 Lunghezza delle foglie

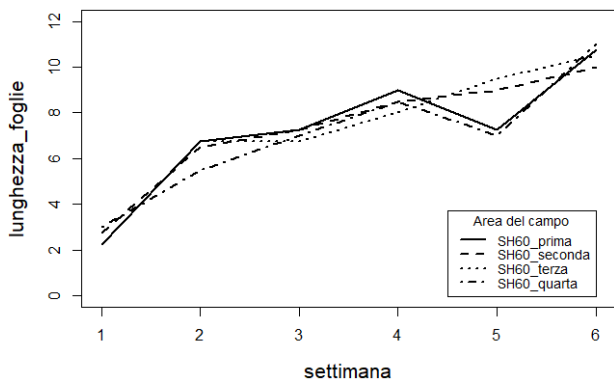
**FL**



**SH30**



**SH60**



**HSH60**

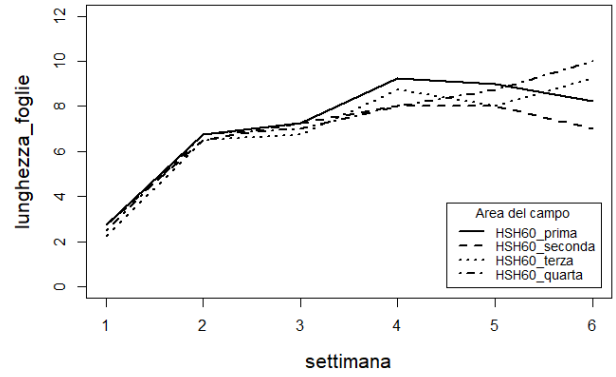
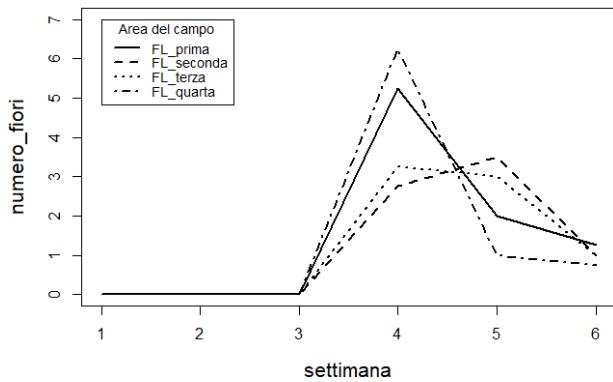


Figura 18a: Curve della lunghezza delle foglie (*length\_leaves*) per ogni area nelle differenti condizioni di luce

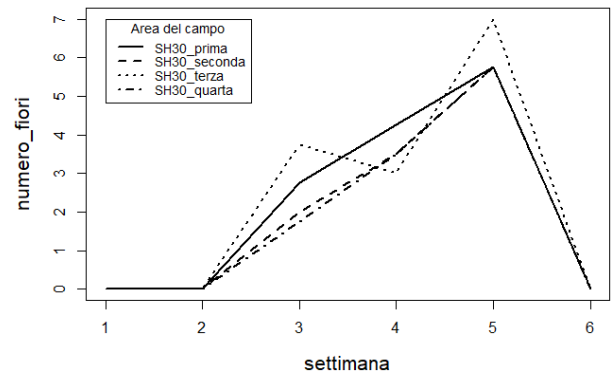
Le curve della lunghezza delle foglie delle piantine nei diversi campi presentano una pendenza inferiore laddove i campi sono influenzati da un impianto agrivoltaico perché l'ombra creata dalla struttura inumidisce maggiormente il terreno e l'aria, rendendo di conseguenza più difficoltoso l'aumento della lunghezza. La fase iniziale però permane molto accentuata anche per i campi non soggetti a luce piena, anche se si scorge una pendenza inferiore per la condizione SH60, situazione soggetta a un'ombreggiatura media costante del 60%.

### 3.2.6 Numero di fiori

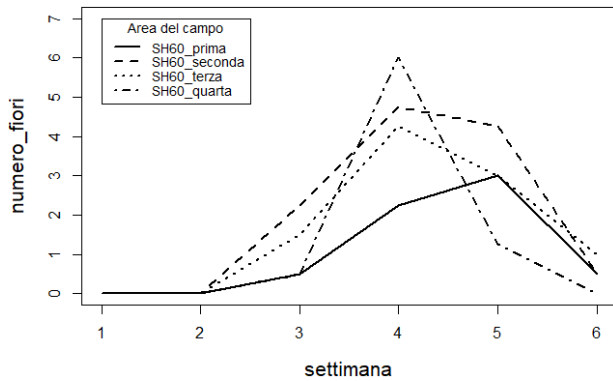
**FL**



**SH30**



**SH60**



**HSH60**

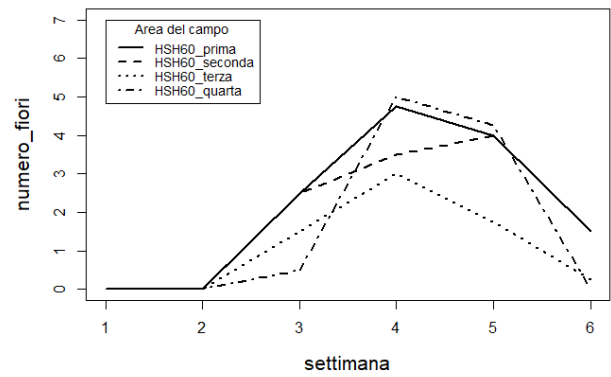


Figura 19a: Curve del numero di fiori (*number\_flowers*) per ogni area nelle differenti condizioni di luce

L'energia solare non favorisce invece la nascita dei fiori, che invece è facilitata per le piantine influenzate dalla presenza di un impianto agrivoltaico; nella condizione di luce piena (FL) i primi fiori iniziano a svilupparsi attorno alla terza settimana, invece per le altre condizioni lo sviluppo è precedente di una settimana.

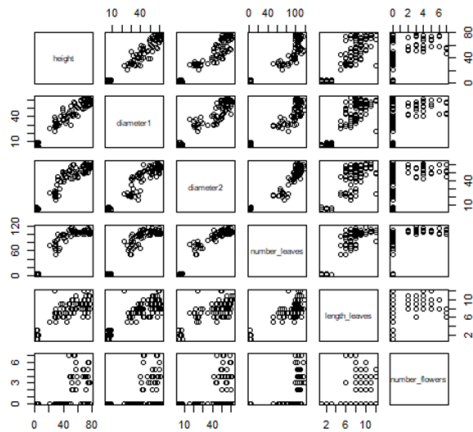
L'andamento del numero di fiori presenta una crescita tanto rapida quanto rapida è la decrescita una volta raggiunto il punto di massimo, ossia la quantità massima di fiori contenuta dalle piantine, soprattutto nelle condizioni FL, SH60 e HSH60.

### 3.3 Condizioni di luce confrontate tra loro (Analisi Bivariata)

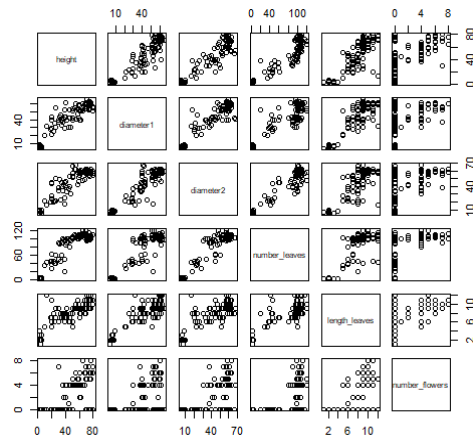
#### 3.3.1 Correlazione delle variabili nelle differenti condizioni di luce

Un aspetto interessante da verificare è la correlazione delle variabili nelle quattro condizioni, sempre considerando il dataset suddiviso in base all'impianto agrivoltaico (FL, SH30, SH60 e HSH60).

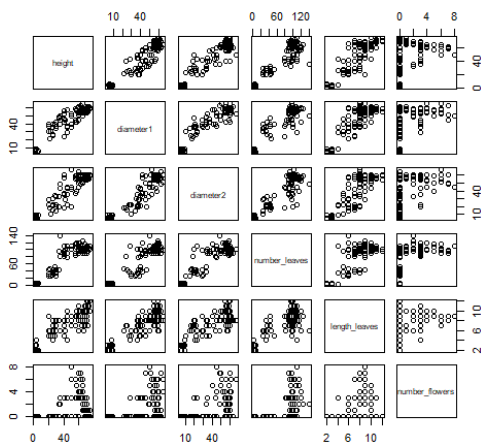
**FL**



**SH30**



**SH60**



**HSH60**

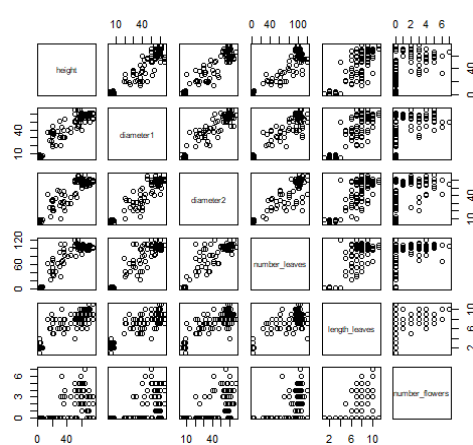


Figura 20a: Correlazioni delle variabili nelle differenti condizioni di luce

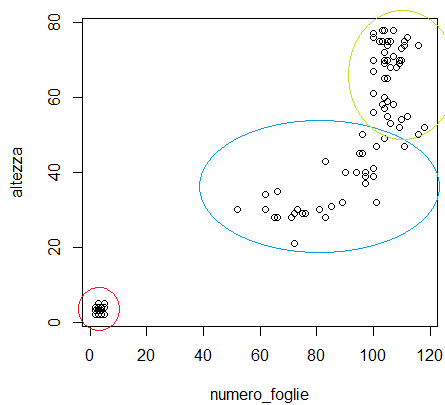
La variabile *height* è fortemente correlata alle variabili che riguardano il diametro (*diameter1* e *diameter2*), oltre che al numero di foglie (*number\_leaves*) e alla loro lunghezza (*length\_leaves*). Tale caratteristica delle foglie risulta leggermente più correlata all'altezza nel caso SH30 e in quello SH60. I due diametri sottoposti a verifica sono correlati tra loro in tutte le situazioni, aspetto inevitabile poiché la crescita risulta proporzionata. Dai risultati si percepisce anche una lieve relazione tra il numero di foglie e il numero di fiori con una correlazione poco marcata.

### 3.3.2 Altezza (height) influenzata dal numero di foglie (number\_leaves)

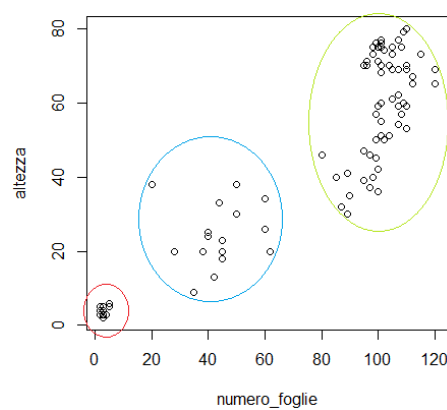
Le foglie svolgono una delle funzioni principali per la sopravvivenza delle piante; esse sono in grado di compiere la fotosintesi clorofilliana, che consiste nel produrre sostanze organiche, la traspirazione e la respirazione, servendosi dell'energia luminosa. I tre processi regolano l'umidità e rinfrescano la temperatura, condizioni necessarie per permettere alla pianta generare ossigeno e gli zuccheri necessari per il suo nutrimento.

Le foglie risultano essere fondamentali per le piante; si vuole a tal proposito cogliere la relazione tra il numero delle foglie nelle differenti condizioni di luce e la crescita delle piantine. Viene proposto un diagramma di dispersione come base per improntare l'analisi, il quale presenta l'altezza e il numero di foglie della piantina per ciascuna condizione di luce.

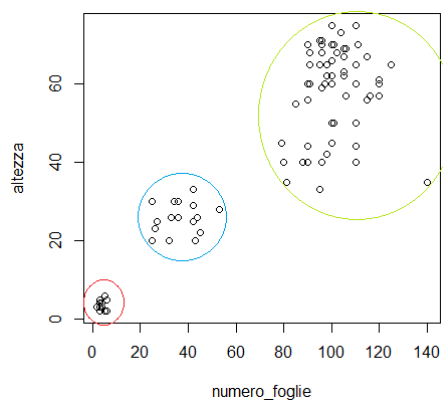
**FL**



**SH30**



**SH60**



**HSH60**

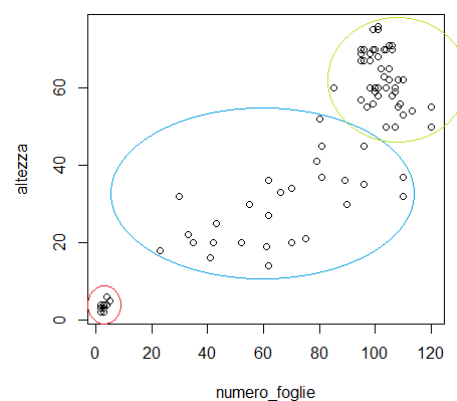


Figura 21a: Grafici di dispersione di height e di number\_leaves nelle differenti condizioni

Dai grafici di dispersione riportati emergono visivamente tre gruppi di dati e una prima osservazione interessa la concentrazione di dati; forte in prossimità di valori molto piccoli, ancora intensa per valori molto elevati e più lieve per i valori centrali. I grafici fanno intendere una dipendenza lineare tra l'altezza della piantina e il suo numero di foglie.

Si stimano i modelli di regressione lineare con l'intercetta ( $\beta_0$ ) e il coefficiente angolare ( $\beta_1$ ) e si riportano le loro stime; si affiancano inoltre i relativi grafici di dispersione con la retta stimata.

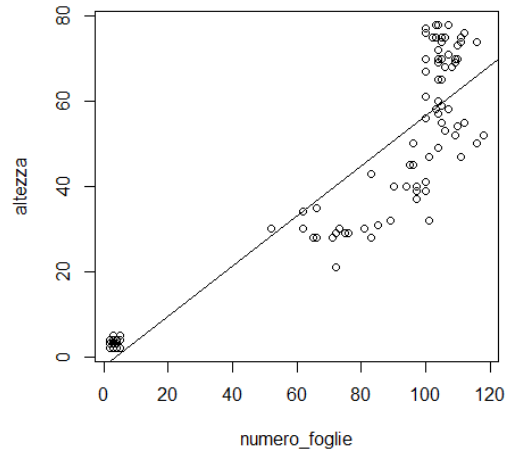
### FL

```
Call:
lm(formula = height ~ number_leaves)

Residuals:
    Min       1Q   Median       3Q      Max
-25.143  -9.966   1.781   8.066  20.445

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.24461    2.79778  -0.802   0.424
number_leaves  0.58799    0.03136  18.751 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.51 on 94 degrees of freedom
Multiple R-squared:  0.789,    Adjusted R-squared:  0.7868
F-statistic: 351.6 on 1 and 94 DF,  p-value: < 2.2e-16
```



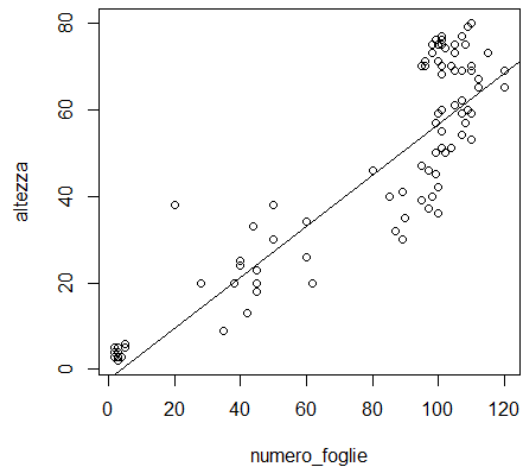
### SH30

```
Call:
lm(formula = height ~ number_leaves)

Residuals:
    Min       1Q   Median       3Q      Max
-23.6532  -7.8021   0.1737   7.9226  25.6362

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.54143    2.38886   0.227   0.821
number_leaves  0.59112    0.02789  21.193 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.81 on 94 degrees of freedom
Multiple R-squared:  0.8269,    Adjusted R-squared:  0.8251
F-statistic: 449.1 on 1 and 94 DF,  p-value: < 2.2e-16
```



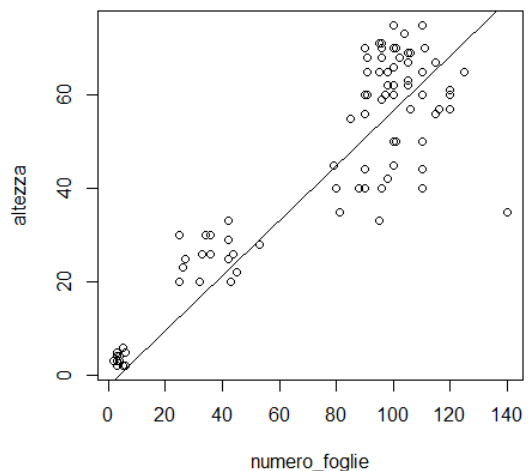
### SH60

```
Call:
lm(formula = height ~ number_leaves)

Residuals:
    Min       1Q   Median       3Q      Max
-44.451  -6.156   1.034   7.714  17.190

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.85783    2.21673   2.191  0.0309 *
number_leaves  0.53281    0.02629  20.266 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.4 on 94 degrees of freedom
Multiple R-squared:  0.8138,    Adjusted R-squared:  0.8118
F-statistic: 410.7 on 1 and 94 DF,  p-value: < 2.2e-16
```





## HSH60

```
Call:
lm(formula = height ~ number_leaves)

Residuals:
    Min       1Q   Median       3Q      Max
-30.544  -5.907   0.643   5.579  18.687

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.2289     2.3523   0.097   0.923
number_leaves 0.5665     0.0279  20.306 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.6 on 94 degrees of freedom
Multiple R-squared:  0.8143,    Adjusted R-squared:  0.8124
F-statistic: 412.3 on 1 and 94 DF,  p-value: < 2.2e-16
```

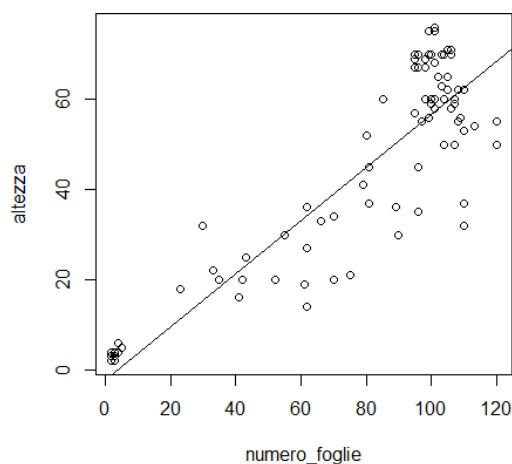


Figura 22a: Output dei modelli stimati e grafico con la retta stimata nelle differenti condizioni

Il coefficiente angolare ( $\beta_1$ ) invece appare significativo in ciascun modello stimato, quindi, a conferma di quanto emerso in precedenza dai grafici di dispersione, esiste una relazione lineare tra l'altezza delle piantine e il numero di foglie. Il parametro inoltre permette di valutare quanto varia mediamente la crescita di una singola piantina quando si incrementa di un'unità il numero di foglie; aggiungendo 10 foglie, la crescita aumenta mediamente di 5.69 cm in FL, di 5.15 cm in SH30, di 5.83 cm in SH60 e di 5.69 cm in HSH60. La condizione di luce con il 30% medio d'ombra vanta una crescita media più elevata, ma in generale le piantine crescono in misura maggiore se influenzate da un impianto agrivoltaico; quindi l'ombreggiatura assiste le foglie consentendoli di limitare l'energia per inumidire e raffreddare il terreno, permettendoli in questo modo di concentrarsi principalmente sul processo chimico per generare zuccheri; dunque producendone una quantità maggiore, la pianta cresce più velocemente.



## 4. MODELLISTICA

Mediante lo studio di vari modelli si vuole cogliere la relazione delle variabili che descrivono le unità statistiche (piantine) nel tempo di osservazione. Per giungere all'obiettivo si confronteranno tra loro il modello di regressione polinomiale rispettivamente per ciascuna variabile con il modello GAM.

### 4.1 Modelli lineari generalizzati (GLM)

I modelli lineari generalizzati GLM corrispondono a una classe che estende il modello di regressione lineare; la famiglia esponenziale si generalizza nella forma:

$$f(y_i; \vartheta_i, \nu) = \exp\left\{\frac{y_i \vartheta_i - c(\vartheta_i)}{\nu} + b(y_i, \nu)\right\}$$

in cui l'argomento della funzione  $\exp()$  integra a sua volta le funzioni note  $b()$  e  $c()$ <sup>7</sup>.

Il modello generalizzato mantiene l'assunzione d'indipendenza ma l'assunzione di linearità e l'assunzione di normalità vengono meno, pertanto dato che  $E[Y_i] < \infty$  si definisce una trasformazione  $g()$  della media  $\mu_i$  invertibile e derivabile:

$$g(E(Y_i)) = \beta^T x_i.$$

La funzione prende il nome di funzione legame (*link function*) ed è espressa dal predittore lineare  $\beta^T x_i$  dove  $x_i = (x_1, \dots, x_p)$  rappresenta un vettore noto delle  $p$  variabili esplicative dell' $i$ -esima osservazione con  $i = 1, \dots, n$  e  $\beta^T = (\beta_1, \dots, \beta_p)^T$  indica il vettore di coefficienti ignoti di regressione. Essa determina la scala sulla quale viene stabilita la linearità nei parametri dell'effetto delle esplicative, ossia garantisce che  $g(\mu_i)$  assuma valori nell'asse reale e non in un intervallo limitato di valori.

Il modello lineare non è altro che un caso particolare del modello generalizzato in quanto della funzione di densità di una normale si considera unicamente l'esponenziale e come funzione di legame si tiene conto della funzione identità per cui  $E[Y_i] = \mu = x_i^T \beta$ <sup>7</sup>.

## VARIABILI RISPOSTA DI CONTEGGIO

Per le variabili risposta discrete che rappresentano un conteggio si assume la distribuzione di Poisson la cui funzione di probabilità può essere espressa in generale:

$$p(y; \vartheta) = \exp(y\vartheta - c(\vartheta) + b(y))$$

la quale costituisce la famiglia esponenziale, dove  $\vartheta$  è un parametro; nello specifico si sviluppa:

$$p(y) = \exp(y \log(\mu) - \mu - \log(y!))$$

essendo  $\mu$  la media, allora  $\vartheta = \log(\mu)$ , ossia  $\mu = e^\vartheta$ .

Quindi  $\vartheta = \log(\mu) = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ ; di conseguenza si avrà  $c(\vartheta) = \mu$  e la funzione canonica sarà  $g(\mu) = \log(\mu)$ <sup>7</sup>.

### 4.2 Modelli di regressione polinomiale

Il modello di regressione polinomiale è un modello multiplo, ossia dispone di più variabili esplicative; di fatto la variabile indipendente è la medesima elevata a differenti gradi, avendo così più variabili a formare il modello.

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + \varepsilon$$

Un modello polinomiale di grado  $k$  assumerà un andamento differente in base al  $k$  fissato, lineare ( $k = 1$ ), quadratico ( $k = 2$ ), cubico ( $k = 3$ ) ecc.

Il coefficiente di determinazione  $R^2$  è la proporzione di variabilità spiegata della variabilità totale che il modello è in grado di spiegare, quindi aumenta ogni qualvolta viene aggiunta una nuova variabile esplicativa; dunque un buon indicatore per definire se un modello è preferibile a uno con un grado inferiore o superiore è il coefficiente di determinazione corretto  $R_c^2$ , in quanto il suo valore aumenta solamente se la nuova variabile apporta effettivamente un miglioramento al modello, quindi solo se rilevanti. La correzione sta nell'introdurre la numerosità campionaria e il numero di regressioni nella formula, dunque considerare la varianza anziché la devianza.

<sup>7</sup>Grigoletto, M., Ventura, L., & Pauli, F. (2017). *Modello Lineare: Teoria e Applicazioni con R*. Giappichelli.

### 4.3 Modelli additivi generalizzati (GAM)

I modelli additivi generalizzati (general additive models, GAM) rientrano nella regressione semi-parametrica in quanto i predittori possono essere modellati non parametricamente utilizzando una funzione liscia. L'assunzione che la variabile risposta sia indipendente con distribuzione di probabilità nota, rende i GAM utili alla modellazione in quanto sono in grado di gestire la non linearità tra la variabile risposta e le esplicative modellando gli effetti non lineari mediante funzioni di lisciamento, pur mantenendo la struttura di tipo additivo classica dei modelli di regressione lineare. Tali funzioni sono applicate a ciascuna variabile esplicativa del modello per riprodurre l'andamento della variabile dipendente al variare delle esplicative e possono dipendere anche da due o più variabili; sono funzioni denominate smoother che utilizzano più polinomi su diverse regioni delle covariate delimitate da estremi detti nodi. La flessibilità dei modelli dipende dal numero di nodi, avendo  $k$  nodi si stimano  $(k + 1)$  polinomi di grado  $q$  dunque sono  $(k + 1)(q + 1)$  i gradi di libertà delle singole funzioni adattate alle variabili esplicative, minori sono i nodi minore è il numero di gradi di libertà<sup>8</sup>.

Un secondo aspetto interessante dei GAM è la capacità di esaminare l'effetto delle variabili esplicative sulla risposta mantenendo fissate le altre variabili, questo perché il modello è additivo, quindi considera le singole funzioni delle variabili esplicative separatamente per poi sommarle.

Il predittore lineare del modello è specificato quanto segue:

$$g(E[Y]) = \beta_0 + \sum_{j=1}^k f_j(x) \beta_j$$

dove  $g$  è la funzione *link* (funzione logaritmica, funzione logit, funzione probit, funzione identità) e  $\sum f_j(x)$  invece è la funzione di lisciamento composta da funzioni più semplici, ossia è l'insieme di funzioni di base ( $f_j(x)$ );  $\beta_0, \beta_1, \dots, \beta_k$  sono i parametri da stimare e rappresentano il peso di ciascuna funzione di base definiti anche penalità<sup>9</sup>.

<sup>8</sup>Bruno, A. A. W. G. S. (2012). *Data Analysis and Data Mining: An Introduction*. Oxford University Press, USA.

<sup>9</sup>Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)* (2nd ed.). Chapman and Hall/CRC.

## INTERPRETAZIONI R

In R, per l'analisi dei modelli GAM esistono due pacchetti utilizzati per la modellazione dei dati.

Il pacchetto **gam** comprende la funzione  $s()$  per spline di lisciamiento univariate e la funzione  $lo()$  per svolgere una regressione locale (local regression). In  $s()$  è possibile includere l'argomento  $df$  per indicare i gradi di libertà (non necessariamente un numero intero) tale che  $df = tr(S) - 1$ , dove  $S$  è la matrice di smoother, e l'argomento  $spar \in (0,1)$ ; entrambi utilizzati come parametro di livellamento. Il termine  $lo()$  (local scatterplot smoother, LOESS) invece è una curva ottenuta applicando la regressione a sottoinsiemi di dati contenuti in intervalli regolari che applica una regressione polinomiale locale assegnando dei pesi a ciascun punto dell'intervallo; pesi maggiori vengono attribuiti a punti più prossimi al centro dell'intervallo. Il grado del polinomio è specificato dell'argomento *degree* (*degree=2* di default) e l'ampiezza è definita dell'argomento *span*  $\in (0,1)$ , che indica la percentuale di dati contenuti nell'intervallo (*span=0.5* di default). Il pacchetto utilizza l'algoritmo di backfitting, il quale prevede la separazione della parte non parametrica da quella parametrica per adattare quest'ultima mediante i minimi quadrati lineari ponderati; si compie la ripetizione dello stesso fino al raggiungimento della convergenza.

Per il procedimento si assume  $\beta_0$  scalare e  $f_i(x) = \hat{\beta}_j x$ , successivamente per ogni  $j = 1, \dots, p$  si considera  $f_j(x) = S(x_j, y - \beta_0 - \sum f_j(x))$ , dove la funzione  $s()$  è scelta dall'utente e può essere uno smoother non parametrico (spline o loess) o parametrico (lineare o polinomiale); il termine al suo interno è un residuo parziale<sup>10</sup>.

Nel pacchetto **mgcv** invece il termine  $s()$  può contenere più di un argomento e implica un lisciatore isotropico, ossia la rotazione di ciascun predittore non modifica lo smoother, il termine  $te()$  modella le interazioni tra variabili esplicative mediante il prodotto tensoriale, costruito da qualsiasi smooth marginale penalizzato singolarmente, i termini  $ti()$  e  $t2()$  trattano rispettivamente le interazioni del prodotto tensoriale con smooth marginali e la costruzione liscia del prodotto in un modo alternativo<sup>9</sup>.

I parametri di smooth sono selezionati mediante le spline di regressione penalizzate che suddividono i dati ordinati in intervalli regolari attraverso dei nodi, quindi le loro stime si ottengono con il metodo dei minimi quadrati penalizzati (OLS), basato sulla massimizzazione della log-verosimiglianza penalizzata:

$$\log(L(\beta)) - \sum \lambda_j \beta_j^T S_j \beta_j$$

caratterizzata da una penalità associata a ciascuna funzione di base definita da  $\int (f_j''(x))^2 dx$  ( $f_j''(x)$  è la derivata seconda) scritta anche in forma matriciale come  $\beta_j^T S_j \beta_j$ ; la penalità serve per disincentivare le oscillazioni. All'aumentare di  $f''$  maggiore sarà la non linearità dello smoother, infatti se  $f'' = 0$  si otterrà una retta. Il termine  $\lambda_j$  invece è definito parametro di lisciamiento (smoothing parameters) e monitora la quantità di livellamento per ciascuna variabile esplicativa, è una sorta di punto di compromesso tra levigatezza del modello stimato e fedeltà dei dati e può essere determinato con una verifica d'ipotesi riferita al numero di nodi; minore è il valore attribuito a  $\lambda_j$ , maggiore sarà la non linearità dello smoother. Entrambi i termini posso essere analizzati con l'ausilio di GCV (Generalized Cross-Validation) per coglierne quello ottimale<sup>10</sup>.

Nel pacchetto gam, per ogni modello GAM occorre definire opportunamente un parametro di regolazione: per la funzione  $s()$  si determina osservando i gradi di libertà ( $df$ ) oppure tramite *spar*, invece per la funzione  $lo()$ , tale parametro di lisciamiento è ottenuto da *span*.

<sup>10</sup>Faraway, J. J. (2005). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. CRC Press.

#### 4.4 Applicazione dei modelli per ciascuna variabile

L'analisi vuole giungere a conclusioni per le variabili risposta *height*, *diameter1*, *diameter2*, *number\_leaves*, *length\_leaves* e *number\_flowers*; si include nei modelli la variabile *week* come esplicativa quantitativa indicatrice del tempo e la variabile *light* mediante dummy che indicano le differenti condizioni di luce che influenzano i campi sottoposti ad analisi; nello specifico:

$$\begin{aligned} x_{i1} = \text{condizione FL} &= \begin{cases} 1 \text{ se FL} \\ 0 \text{ altrimenti} \end{cases} & x_{i3} = \text{condizione SH60} &= \begin{cases} 1 \text{ se SH60} \\ 0 \text{ altrimenti} \end{cases} \\ x_{i2} = \text{condizione SH30} &= \begin{cases} 1 \text{ se SH30} \\ 0 \text{ altrimenti} \end{cases} & x_{i4} = \text{condizione HSH60} &= \begin{cases} 1 \text{ se HSH60} \\ 0 \text{ altrimenti} \end{cases} \end{aligned}$$

$x_{i1}$ , ossia la dummy riferita alla condizione di luce FL, viene utilizzata come riferimento.

Per ciascuna variabile dipendente si riporta il modello polinomiale e il modello gam considerati più opportuni, ognuno espresso in funzione della variabile *week* e della variabile *light*. La selezione del modello di regressione polinomiale avviene in seguito a un'analisi di diversi modelli di regressione lineare di secondo e di terzo ordine i quali contengono anche l'interazione tra la variabile *week* e la variabile *light*; anche se i coefficienti d'interazione non risultano significativi, si opta per includerli comunque nel modello, in quanto l'obiettivo è quello di stimare una curva di crescita per ogni condizione di luce. Il modello polinomiale occorre per introdurre la flessibilità ma ci si limita al grado 3 in quanto con un grado più elevato si avrebbe maggior flessibilità; per questa si preferisce adattare ai dati un modello additivo generalizzato (GAM). Per determinare il modello GAM invece si utilizza la funzione **gam** dell'omonimo pacchetto con la funzione di lisciamento  $s()$ , la quale contiene anche il termine d'interazione esprimibile nel modello introducendo le variabili in interazione tra loro separate da virgola ( $f(\text{variabile}_1, \text{variabile}_2)$ ). Il modello GAM considerato preferibile è stato identificato tra diversi modelli che differiscono tra loro per il grado di lisciamento. Si nota inoltre che non esistono differenze significative fra le diverse funzioni di splines se non nel parametro di lisciamento, che però richiederebbe un approccio diverso da quello trattato, per cui vengono usati i parametri impostati di default.



### 4.4.1 Altezza

```
Call:
lm(formula = height ~ poly(week, 2) * light, data = altezza_dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-21.0357  -3.5286   0.0458   3.0982  15.3929

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  45.36458    0.53007   85.582 < 2e-16 ***
poly(week, 2)1  465.84035    10.38727   44.847 < 2e-16 ***
poly(week, 2)2 -96.17954    10.38727   -9.259 < 2e-16 ***
lightHSH60    -2.71875     0.74964   -3.627 0.000327 ***
lightSH30      0.08333     0.74964    0.111 0.911545
lightSH60     -1.06250     0.74964   -1.417 0.157216
poly(week, 2)1:lightHSH60 -11.65348    14.68981   -0.793 0.428107
poly(week, 2)2:lightHSH60  1.30931    14.68981    0.089 0.929026
poly(week, 2)1:lightSH30  16.01606    14.68981    1.090 0.276294
poly(week, 2)2:lightSH30  -0.60010    14.68981   -0.041 0.967436
poly(week, 2)1:lightSH60 -17.92843    14.68981   -1.220 0.223061
poly(week, 2)2:lightSH60 -12.11109    14.68981   -0.824 0.410209
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.194 on 372 degrees of freedom
Multiple R-squared:  0.9572,    Adjusted R-squared:  0.9559
F-statistic: 756.1 on 11 and 372 DF,  p-value: < 2.2e-16
```

```
Call: gam(formula = height ~ s(week, df = 4) * light, data = altezza_dataset)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-19.692  -2.742  -0.314   2.907  14.718

(Dispersion Parameter for gaussian family taken to be 25.3293)

Null Deviance: 234386.6 on 383 degrees of freedom
Residual Deviance: 9447.845 on 373 degrees of freedom
AIC: 2343.658

Number of Local Scoring Iterations: 2

Anova for Parametric Effects
              Df Sum Sq Mean Sq  F value    Pr(>F)
s(week, df = 4)  1 213859  213859 8443.1316 < 2.2e-16 ***
light            3    490    163   6.4542 0.0002864 ***
s(week, df = 4):light  3    167    56   2.1969 0.0880305 .
Residuals       373   9448    25
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects
              Npar Df Npar F      Pr(F)
(Intercept)
s(week, df = 4)  3 137.16 < 2.2e-16 ***
light
s(week, df = 4):light
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 1b: Output in R del modello polinomiale a sx e del modello GAM a dx per la variabile height

### GRAFICI

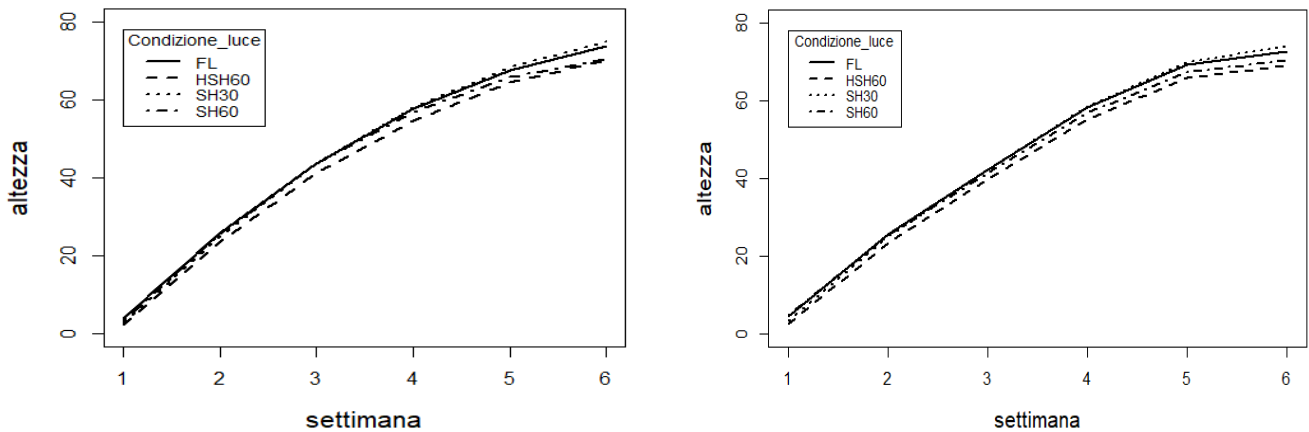


Figura 2b: Grafico del modello polinomiale e del modello GAM per la variabile height

Il modello polinomiale per la risposta riferita all'altezza (*height*) delle piantine ha un grado pari a 2, il modello GAM invece ha un grado di lisciamento pari a 4; entrambi i modelli includono l'interazione. Da un'analisi grafica gli andamenti globali dei due modelli sono simili e le condizioni di luce non differenziano notevolmente tra loro all'interno degli specifici modelli.

## 4.4.2 Diametro Principale

```
Call:
lm(formula = diameter1 ~ poly(week, 2) * light, data = diameter1_dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-21.0768  -3.5010  -0.0893   2.9687  26.5786

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    39.6042     0.6289   62.974 < 2e-16 ***
poly(week, 2)1  325.8193    12.3237   26.438 < 2e-16 ***
poly(week, 2)2  -92.6880    12.3237   -7.521 4.11e-13 ***
lightHSH60      1.7292     0.8894    1.944 0.0526 .
lightSH30       2.1146     0.8894    2.378 0.0179 *
lightSH60       2.7813     0.8894    3.127 0.0019 **
poly(week, 2)1:lightHSH60 11.7132    17.4284    0.672 0.5020
poly(week, 2)2:lightHSH60 -40.1521    17.4284   -2.304 0.0218 *
poly(week, 2)1:lightSH30  2.9283    17.4284    0.168 0.8667
poly(week, 2)2:lightSH30 -12.7112    17.4284   -0.729 0.4663
poly(week, 2)1:lightSH60  -4.3626    17.4284   -0.250 0.8025
poly(week, 2)2:lightSH60 -44.6801    17.4284   -2.564 0.0108 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.162 on 372 degrees of freedom
Multiple R-squared:  0.8965, Adjusted R-squared:  0.8934
F-statistic: 292.9 on 11 and 372 DF,  p-value: < 2.2e-16

Call:
gam(formula = diameter1 ~ s(week, df = 2) * light, data = diameter1_dataset)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-18.2738  -4.5180  -0.9815   3.0688  26.7125

(Dispersion Parameter for gaussian family taken to be 41.8815)

Null Deviance: 136458 on 383 degrees of freedom
Residual Deviance: 15705.56 on 374.9999 degrees of freedom
AIC: 2534.818

Number of Local Scoring Iterations: 2

Anova for Parametric Effects
            Df Sum Sq Mean Sq F value Pr(>F)
s(week, df = 2)  1 107839  107839 2574.8686 < 2e-16 ***
light           3    406    135   3.2275 0.02257 *
s(week, df = 2):light  3    35    12   0.2754 0.84317
Residuals      375 15706    42
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects
            Npar Df Npar F      Pr(F)
(Intercept)
s(week, df = 2)  1 297.79 < 2.2e-16 ***
light
s(week, df = 2):light
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 3b: Output in R del modello polinomiale a sx e del modello GAM a dx per la variabile diameter1

### GRAFICI

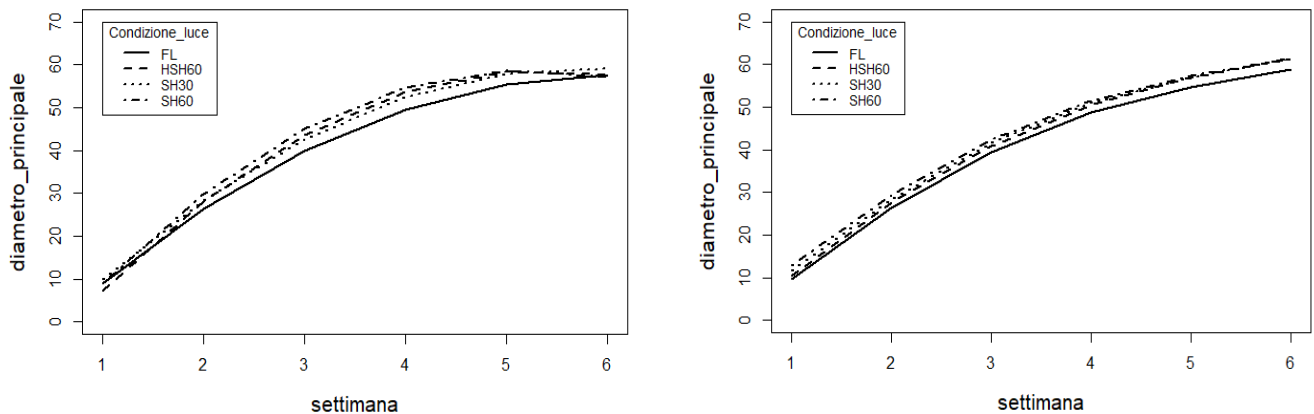


Figura 4b: Grafico del modello polinomiale e del modello GAM per la variabile diameter1

Il diametro principale individuato dalla variabile risposta diameter1 stima un modello polinomiale di ordine 2 e un modello GAM con grado di lisciamento pari a 2.

Le curve nei modelli appaiono simili e in entrambi la condizione di luce piena (FL) assume valori inferiori rispetto a quelli assunti dalle condizioni di luce influenzate da un impianto agrivoltaico; nel modello polinomiale tale affermazione la si può trarre dalla seconda settimana di osservazione, mentre nel modello GAM già dalla prima settimana. In generale le condizioni di luce soggette da impianto non differenziano tra loro ma solamente dalla condizione di luce piena FL nella fase intermedia dell'osservazione.

### 4.4.3 Diametro Secondario

```

Call:
lm(formula = diameter2 ~ poly(week, 2) * light, data = diameter2_dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-22.1750  -3.1161  -0.0067   2.9029  21.8906

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    39.3542    0.6493   60.614 < 2e-16 ***
poly(week, 2)1  338.3692   12.7229   26.595 < 2e-16 ***
poly(week, 2)2 -131.4763   12.7229  -10.334 < 2e-16 ***
lightHSH60      1.6771     0.9182    1.827  0.06858 .
lightSH30       2.6250     0.9182    2.859  0.00449 **
lightSH60       0.1563     0.9182    0.170  0.86497
poly(week, 2)1:lightHSH60 12.3706   17.9929    0.688  0.49218
poly(week, 2)2:lightHSH60 14.4569   17.9929    0.803  0.42221
poly(week, 2)1:lightSH30  11.8328   17.9929    0.658  0.51118
poly(week, 2)2:lightSH30 -2.4550   17.9929   -0.136  0.89155
poly(week, 2)1:lightSH60  15.7173   17.9929    0.874  0.38294
poly(week, 2)2:lightSH60  23.2402   17.9929    1.292  0.19729
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.361 on 372 degrees of freedom
Multiple R-squared:  0.901, Adjusted R-squared:  0.8981
F-statistic: 307.8 on 11 and 372 DF, p-value: < 2.2e-16

Call:
gam(formula = diameter2 ~ s(week, df = 3) * light, data = diameter2_dataset)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-21.6992  -2.9038  -0.6024   2.9119  23.1959

(Dispersion Parameter for gaussian family taken to be 39.4901)

Null Deviance: 152051.6 on 383 degrees of freedom
Residual Deviance: 14769.31 on 374.0001 degrees of freedom
AIC: 2513.216

Number of Local Scoring Iterations: 2

Anova for Parametric Effects
              Df Sum Sq Mean Sq F value Pr(>F)
s(week, df = 3)  1 121347  121347 3072.8519 < 2.2e-16 ***
light            3    457    152    3.8559  0.009714 **
s(week, df = 3):light  3    35    12    0.2989  0.826162
Residuals      374 14769    39
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects
              Npar DF Npar F      Pr(F)
(Intercept)
s(week, df = 3)  2 195.53 < 2.2e-16 ***
light
s(week, df = 3):light
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 5b: Output in R del modello polinomiale a sx e del modello GAM a dx per la variabile diameter2

### GRAFICI

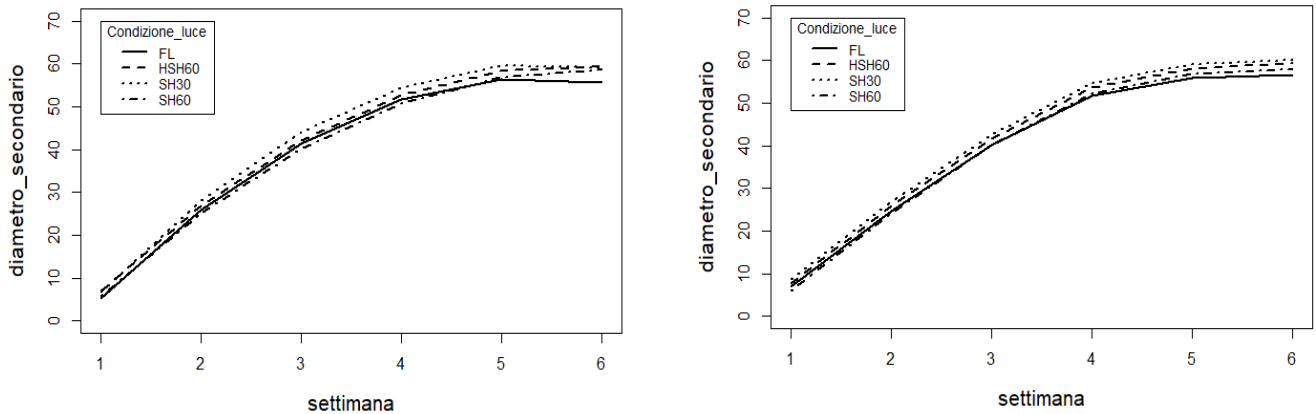


Figura 6b: Grafico del modello polinomiale e del modello GAM per la variabile diameter2

Il diametro secondario definisce un modello di regressione polinomiale con un grado pari a 2, e un modello GAM con un grado di lisciamiento pari a 3.

Nel modello polinomiale la curva riferita alla condizione FL sembra diminuire più rapidamente rispetto alle curve riferite alle condizioni di luce artificiali, le quali tra loro non sembrano adottare un andamento diverso. Nel modello GAM invece gli andamenti sono uguali per tutte le condizioni di luce e FL assume valori inferiori rispetto a quelli assunti dalle rimanenti condizioni di luce soprattutto dopo la quarta settimana di rilevazione.

## 4.4.4 Numero di foglie

```
Call:
glm(formula = number_leaves ~ poly(week, 3) * light, family = poisson,
     data = number_leaves_dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4052  -0.9500   0.2366   0.6222   4.7674

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.12264    0.01687  244.339 < 2e-16 ***
poly(week, 3)1  14.03655    0.41565   33.770 < 2e-16 ***
poly(week, 3)2 -10.56237    0.36465  -28.966 < 2e-16 ***
poly(week, 3)3   5.23993    0.28724   18.242 < 2e-16 ***
lightHSH60    -0.15335    0.02595   -5.910 3.42e-09 ***
lightSH30    -0.15423    0.02618   -5.892 3.82e-09 ***
lightSH60    -0.20815    0.02708   -7.687 1.51e-14 ***
poly(week, 3)1:lightHSH60  2.71552    0.64711   4.196 2.71e-05 ***
poly(week, 3)2:lightHSH60 -1.53114    0.56060   -2.731 0.006310 **
poly(week, 3)3:lightHSH60 -0.20890    0.42768   -0.488 0.625233
poly(week, 3)1:lightSH30  3.28387    0.65430   5.019 5.20e-07 ***
poly(week, 3)2:lightSH30 -1.75374    0.56512   -3.103 0.001914 **
poly(week, 3)3:lightSH30 -0.46293    0.42807   -1.081 0.279495
poly(week, 3)1:lightSH60  4.22823    0.67933   6.224 4.84e-10 ***
poly(week, 3)2:lightSH60 -2.25365    0.58439   -3.856 0.000115 ***
poly(week, 3)3:lightSH60 -0.21120    0.43724   -0.483 0.629079
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 11208.45 on 383 degrees of freedom
Residual deviance: 632.56 on 368 degrees of freedom
AIC: 2875.9

Number of Fisher Scoring iterations: 5
```

```
Call:
gam(formula = number_leaves ~ s(week, df = 5) * light, family = poisson,
     data = number_leaves_dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.49260  -0.58195  -0.07318   0.53123   5.10358

(Dispersion Parameter for poisson family taken to be 1)

Null deviance: 11208.45 on 383 degrees of freedom
Residual deviance: 455.1201 on 372 degrees of freedom
AIC: 2690.486

Number of Local Scoring Iterations: 4

Anova for Parametric Effects

            Df Sum Sq Mean Sq F value    Pr(>F)
s(week, df = 5)  1 1273.62 1273.62 1046.080 < 2.2e-16 ***
light           3   37.51   12.50  10.269 1.645e-06 ***
s(week, df = 5):light  3  42.22   14.07  11.559 2.925e-07 ***
Residuals      372  452.92    1.22

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects

(Intercept)  Npar Df Npar chisq    P(Chi)
s(week, df = 5)      4    2586.3 < 2.2e-16 ***
light
s(week, df = 5):light
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 7b: Output in R del modello polinomiale a sx e del modello GAM a dx per la variabile number\_leaves

### GRAFICI

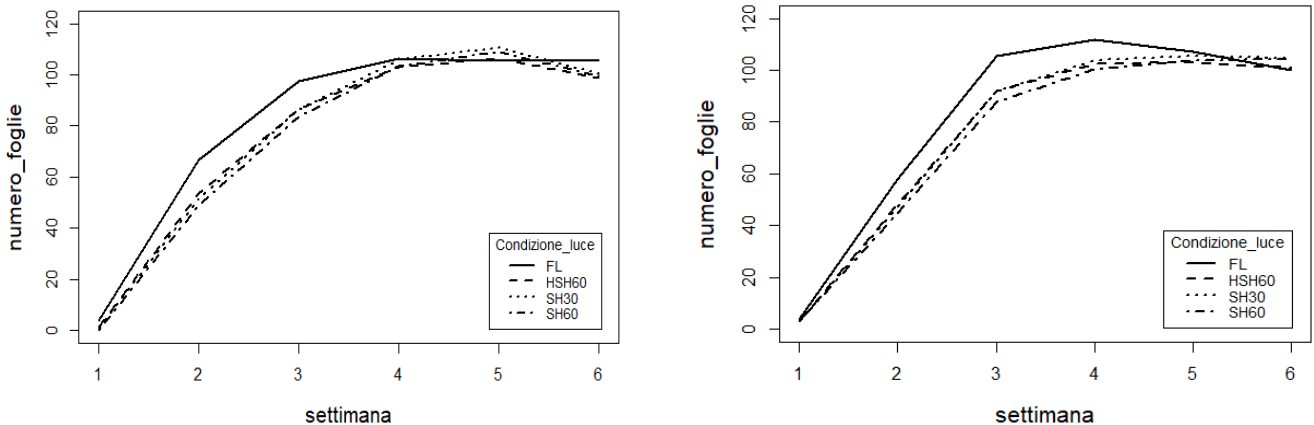


Figura 8b: Grafico del modello polinomiale e del modello GAM per la variabile number\_leaves

Il numero di foglie è una variabile di conteggio, dunque si utilizza il modello lineare generalizzato. Il modello polinomiale stimato di ordine 3 sembra il più adeguato, mentre per il modello GAM il grado di lisciamento tende ad essere preferito pari a 5.

Nel modello polinomiale la curva FL si discosta dalle altre curve fino alla quarta settimana, mentre le curve delle condizioni con agrivoltaico presentano un andamento simile tra loro. Nel GAM, FL mostra un andamento superiore rispetto a quello delle altre condizioni fino alla quinta settimana in cui si avvicina alle altre condizioni, per le quali permangono andamenti simili tra loro.

## 4.4.5 Lunghezza delle foglie

```
Call:
lm(formula = length_leaves ~ poly(week, 3) * light, data = length_leaves_dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-3.7192 -0.7059 -0.0913  0.7460  5.9469

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)      7.0729     0.1348   52.483 < 2e-16 ***
poly(week, 3)1    38.7852     2.6409   14.687 < 2e-16 ***
poly(week, 3)2   -23.4584     2.6409   -8.883 < 2e-16 ***
poly(week, 3)3     6.1119     2.6409    2.314 0.021197 *
lightHSH60       -0.1042     0.1906   -0.547 0.585014
lightSH30         0.6146     0.1906    3.225 0.001374 **
lightSH60         0.1354     0.1906    0.711 0.477829
poly(week, 3)1:lightHSH60 -3.2271     3.7347   -0.864 0.388105
poly(week, 3)2:lightHSH60  4.8008     3.7347    1.285 0.199444
poly(week, 3)3:lightHSH60  0.9690     3.7347    0.259 0.795436
poly(week, 3)1:lightSH30  9.3825     3.7347    2.512 0.012424 *
poly(week, 3)2:lightSH30  7.7467     3.7347    2.074 0.038751 *
poly(week, 3)3:lightSH30  4.1740     3.7347    1.118 0.264460
poly(week, 3)1:lightSH60  6.0359     3.7347    1.616 0.106917
poly(week, 3)2:lightSH60 13.6932     3.7347    3.666 0.000282 ***
poly(week, 3)3:lightSH60  6.7455     3.7347    1.806 0.071712 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.32 on 368 degrees of freedom
Multiple R-squared:  0.7745,    Adjusted R-squared:  0.7653
F-statistic: 84.25 on 15 and 368 DF,  p-value: < 2.2e-16
```

```
Call: gam(formula = length_leaves ~ s(week, df = 4) * light, data = length_leaves_dataset)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3757 -0.7481 -0.1189  0.7254  5.5794

(Dispersion Parameter for gaussian family taken to be 1.6693)

Null Deviance: 2844.906 on 383 degrees of freedom
Residual Deviance: 622.6472 on 373 degrees of freedom
AIC: 1299.346

Number of Local Scoring Iterations: 2

Anova for Parametric Effects
              Df Sum Sq Mean Sq F value Pr(>F)
s(week, df = 4) 1 1750.00 1750.00 1048.3465 < 2.2e-16 ***
light           3   29.05    9.68   5.8013 0.0006959 ***
s(week, df = 4):light 3 24.43    8.14  4.8784 0.0024373 **
Residuals     373  622.65    1.67

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects
              Npar Df Npar F      Pr(F)
(Intercept)          3  83.624 < 2.2e-16 ***
s(week, df = 4)          3  83.624 < 2.2e-16 ***
light                    3  83.624 < 2.2e-16 ***
s(week, df = 4):light    3  83.624 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 9b: Output in R del modello polinomiale a sx e del modello GAM a dx per la variabile length\_leaves

### GRAFICI

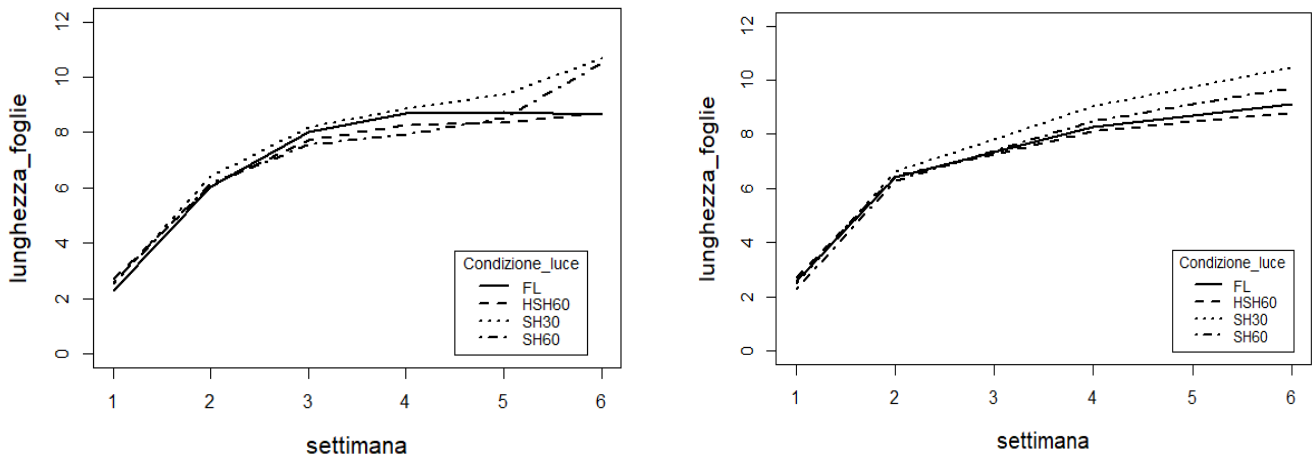


Figura 10b: Grafico del modello polinomiale e del modello GAM per la variabile length\_leaves

Per la lunghezza delle foglie il modello polinomiale più opportuno è quello cubico; il GAM invece appare più efficace se la funzione di lisciamento  $s()$  ha gradi di libertà ( $df$ ) pari a 4. Nella prima settimana di osservazione, il polinomiale conduce a curve simili per tutte le condizioni di luce trattate, in seguito SH30 e SH60 si differenziano da FL e HSH60 aumentando più rapidamente. Anche il GAM riporta curve simili nella prima settimana, successivamente SH30 si discosta e assume valori maggiori rispetto alle rimanenti. Le condizioni di luce globalmente hanno un andamento crescente.

## 4.4.6 Numero di fiori

```
Call:
glm(formula = number_flowers ~ poly(week, 2) * light, family = poisson,
     data = n_flowers_dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3801 -0.8791 -0.1340  0.2575  3.5499

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    -2.4938    0.4879  -5.111 3.20e-07 ***
poly(week, 2)1  69.7620   10.1142   6.897 5.29e-12 ***
poly(week, 2)2 -47.6728    6.4215  -7.424 1.14e-13 ***
lightHSH60      1.6351    0.5488   2.979 0.00289 **
lightSH30       1.5171    0.5533   2.742 0.00611 **
lightSH60       0.9647    0.5905   1.634 0.10231
poly(week, 2)1:lightHSH60 -32.3950   11.3992  -2.842 0.00449 **
poly(week, 2)2:lightHSH60  15.3860    7.5262   2.044 0.04092 *
poly(week, 2)1:lightSH30 -28.1032   11.3986  -2.465 0.01368 *
poly(week, 2)2:lightSH30   9.2344    7.6082   1.214 0.22485
poly(week, 2)1:lightSH60 -22.7919   12.0395  -1.893 0.05835 .
poly(week, 2)2:lightSH60   3.7658    8.2288   0.458 0.64721
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 1225.45 on 383 degrees of freedom
Residual deviance: 413.86 on 372 degrees of freedom
AIC: 947.66

Number of Fisher Scoring iterations: 7
```

```
Call: gam(formula = number_flowers ~ s(week, df = 2) * light, family = poisson,
          data = n_flowers_dataset)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9724 -1.3033 -0.4108  0.2998  2.7976

(Dispersion Parameter for poisson family taken to be 1)

Null Deviance: 1225.448 on 383 degrees of freedom
Residual Deviance: 516.1182 on 375.0001 degrees of freedom
AIC: 1043.916

Number of Local Scoring Iterations: 15

Anova for Parametric Effects
              Df Sum Sq Mean Sq F value    Pr(>F)
s(week, df = 2) 1  25.58  25.5755  23.9267 1.488e-06 ***
light           3  14.56   4.8520   4.5392 0.003858 **
s(week, df = 2):light 3   6.70   2.2348   2.0907 0.101007
Residuals     375 400.84   1.0689
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects
              Npar Df Npar Chisq    P(Chi)
(Intercept)           1
s(week, df = 2)       1  451.33 < 2.2e-16 ***
light
s(week, df = 2):light
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 11b: Output in R del modello polinomiale a sx e del modello GAM a dx per la variabile number\_flowers

### GRAFICI

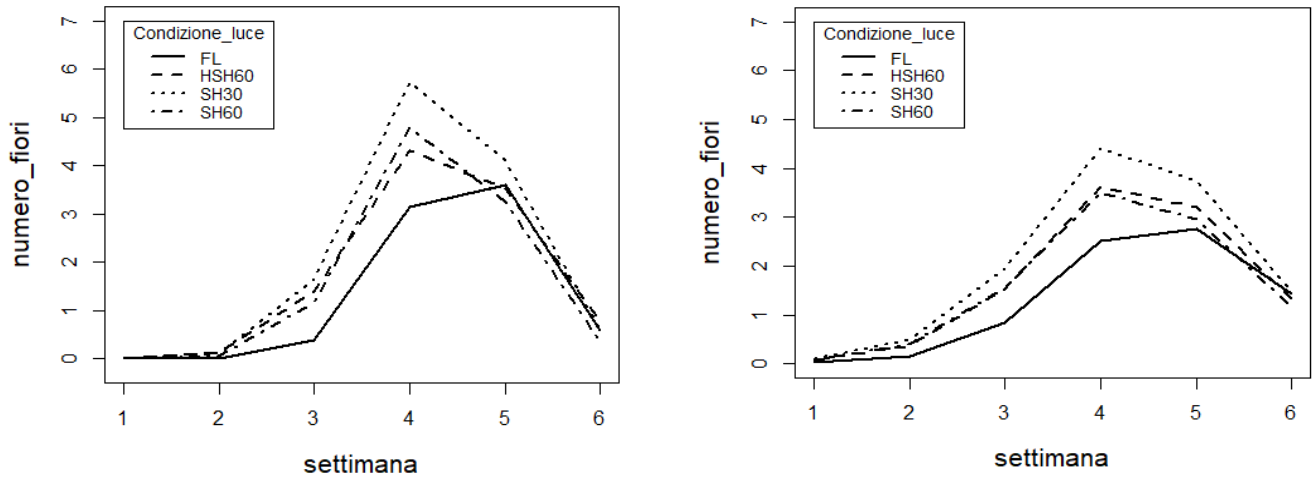


Figura 12b: Grafico del modello polinomiale e del modello GAM per la variabile number\_flowers

Il polinomiale viene stimato con un `glm()` di ordine 2, il gam invece con il grado pari a 2. Nel polinomiale HSH60 e SH60 riportano un andamento simile, mentre SH30 e FL assumono valori inferiori ma simili tra loro nella prima settimana; FL continua ad assumere valori inferiori mentre SH30 assume valori maggiori dalla terza fino alla quinta settimana, in cui tutte le curve delle condizioni di luce diminuiscono. Il modello GAM presenta la medesima struttura del polinomiale anche se ciascuna condizione di luce assume valori inferiori rispetto a quelli assunti nel polinomiale, inoltre la crescita nei GAM è meno rapida attorno alla terza settimana.

## COMMENTO

Il modello polinomiale occorre per introdurre la flessibilità ma ci si limita al grado 3 in quanto con un grado più elevato si avrebbe maggior flessibilità; per questa si preferisce adattare ai dati un modello additivo generalizzato (*GAM*). L'analisi vede interessato il confronto tra i due modelli per ciascuna variabile risposta trattata singolarmente, l'obiettivo è quello di stabilire se il *GAM* stima una relazione più fedele ai dati osservati, dato che è in grado di cogliere una maggiore flessibilità. Dalle analisi condotte si conclude che non emergono particolari differenze negli andamenti delle curve delle condizioni nei due modelli.

Osservando le singole variabili risposta analizzate, in riferimento a ciascuna condizioni di luce, il modello polinomiale stimato non differenzia dal modello *GAM* se si considera l'altezza e i due diametri delle piantine. Per la variabile risposta che rappresenta il numero di foglie invece i due modelli trattati mostrano sottili differenze tra loro e globalmente le condizioni presentano un andamento simile all'interno dei corrispettivi modelli. In entrambi la condizione di luce piena FL assume valori maggiori rispetto a quelli assunti dalle condizioni influenzate da un impianto, anche se nel *GAM* i valori appaiono maggiori rispetto che nel polinomiale, inoltre la rapida crescita delle curve nei *GAM* permane fino alla terza settimana per poi decrescere dalla quinta, mentre nel polinomiale la crescita generale rallenta una settimana in anticipo per poi stabilizzarsi attorno alla quarta settimana. Trattando la lunghezza delle foglie, inizialmente gli andamenti del polinomiale sono simili a quelli del *GAM*, successivamente appaiono delle differenze, in quanto nelle settimane intermedie le curve del polinomiale riportano valori inferiori e nella fase conclusiva dell'osservazione le curve SH30 e SH60 intraprendono un andamento molto differente da quello assunto nel *GAM*.

Il modello polinomiale per il numero di fiori riporta una rapida crescita attorno alla terza settimana e una rapida decrescita dalla quinta, assumendo valori nel mentre più elevati rispetto a quelli assunto nella medesima fascia di osservazione dalle curve dei modelli *GAM*, i quali crescono e decrescono in modo analogo ma meno bruscamente.





## 5. DIFFERENZA TRA LE CONDIZIONI DI LUCE

Il test *ANOVA* (*Analysis of Variance*) ha come obiettivo quello di valutare gli effetti su una variabile risposta d'interesse di uno o più fattori che rappresentano le variabili esplicative categoriali, nello specifico s'intende definire se le differenze in istanti diversi sono uguali in condizioni di luci diverse, dunque stabilire gli effetti delle variabili categoriali con una o più modalità sulla variabile dipendente.

Per svolgere il test è indispensabile verificare la normalità della distribuzione della variabile dipendente, ad esempio tramite il test di Shapiro-Wilk, e l'omoschedasticità, ossia l'omogeneità delle varianze di gruppi tramite il test di Bartlett.

Avendo un'unica variabile dipendente si definisce:

- **Il test *ANOVA* a un fattore**  
per il quale vede interessato un modello contenente una variabile esplicative categoriale con due o più modalità
- **il test *ANOVA* a due fattori**  
che si svolge se il modello si compone di più di una variabile esplicative categoriale, ciascuna con una o più modalità.
- **il test *ANCOVA***  
è una generalizzazione dei due test precedenti in quanto include anche una o più variabili esplicative.

## 5.1 ANOVA a un fattore

Il test *ANOVA* a un fattore si utilizza quando il modello include un'unica variabile esplicativa categoriale con due o più modalità, esso prevede la scomposizione della devianza totale in devianza tra i gruppi, ossia la variabilità tra i gruppi, e in devianza entro i gruppi, ossia la variabilità interna ai gruppi intesa come la dispersione delle osservazioni attorno alla media di gruppo. In particolare sia  $y_{ij}$  la  $i$ -esima osservazione nel  $j$ -esimo gruppo, si assume che le  $y_{ij}$  siano la realizzazione di variabili casuali normali indipendenti e che  $J$  sia il numero di gruppi osservati, allora si ha:

$$\sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 = \sum_{j=1}^J (n_j - 1) s_j^2 + \sum_{j=1}^J n_j (\bar{y}_j - \bar{y})^2$$

Se la variabilità tra i gruppi è elevata rispetto a quella entro i gruppi ( $SQ_{reg} > SQ_{res}$ ) allora esistono differenze significative tra le medie dei gruppi.

Il test mira a verificare l'ipotesi con ipotesi nulla: uguaglianza delle medie dei gruppi, ovvero  $H_0: \mu_1 = \mu_2 = \dots = \mu_j$ , contro l'alternativa in cui almeno una coppia di gruppi abbia una differenza significativa; la statistica test utilizzata si distribuisce come una  $F$  di Fisher con  $(J - 1)$  gradi di libertà al numeratore e  $(n - J)$  al denominatore ed è definita come il rapporto che segue:

$$\frac{SQ_{reg}/(J - 1)}{SQ_{res}/(n - J)} \sim F_{(J-1), (n-J)}$$

Si rifiuta l'ipotesi nulla se la statistica test assume un valore più elevato rispetto al quantile di ordine  $1-\alpha$  della distribuzione  $F$  di Fisher con gradi di libertà pari a  $(J - 1)$  e  $(n - J)$ .

Il test *ANOVA* può essere espresso mediante un modello di regressione lineare multiplo:

$$Y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_J x_{iJ} + \varepsilon_i$$

I parametri  $\beta_2, \dots, \beta_J$  sono le variazioni attese del livello 1 (modalità di riferimento) a livello 2, ...,  $J$  (le rimanenti modalità), di conseguenza l'ipotesi nulla sarà:  $H_0: \beta_2 = \dots = \beta_J = 0$ .

## 5.2 ANOVA a due fattori

L'ANOVA a un fattore può essere esteso a due fattori se il modello contiene due variabili esplicative categoriali ciascuna con due o più modalità. L'obiettivo è verificare l'effetto dei due fattori sulla risposta, ossia stabilire se la distribuzione della variabile dipendente coincide con quella per i diversi livelli, ossia per le diverse modalità delle variabili categoriali. Nello specifico si svolge un'analisi senza interazioni per stabilire l'effetto dei due fattori assumendo che il fattore non dipende da quello a cui ci si condiziona, infine si verifica una possibile intersezione per verificare se l'effetto della variabile qualitativa si modifica al variare del livello della quantitativa a cui ci si sta condizionando.

Per l'ANOVA senza intersezioni il modello ha la seguente struttura:

$$Y_i = \mu + \alpha_2 x_{i,2} + \dots + \alpha_J x_{i,J} + \gamma_2 x_{i,J+1} + \dots + \gamma_K x_{i,J+K-1} + \varepsilon_i$$

Specificando il modello completo e quello ridotto rispettivamente con  $M_1$  e  $M_0$  e il numero di parametri dei due modelli rispettivamente con  $p$  e  $p_0$  si verifica l'assenza dei fattori:

	<b>IPOTESI NULLA</b>	coeff. $M_1$	coeff. $M_0$
<b>primo</b> fattore	$\alpha_2 = \dots = \alpha_J$	$p = J + K - 1$	$p_0 = K$
<b>secondo</b> fattore	$\gamma_2 = \dots = \gamma_K$	$p = J + K - 1$	$p_0 = J$
<b>entrambi</b> fattori	$\alpha_2 = \dots = \alpha_J = \gamma_2 = \dots = \gamma_K$	$p = J + K - 1$	$p_0 = J + K - 2$

Con  $SQ_I$  e  $SQ_{II}$  rispettivamente la devianza dovuta al primo e al secondo fattore e  $SQ_{res}$  la residua, le statistiche test si definiscono:

	Statistica test F	distribuzione
<b>primo</b> fatt.	$SQ_I / (J-1) / SQ_{res} / [n - (J + K - 1)]$	$F_{J-1, n-(J+K-1)}$
<b>secondo</b> fatt.	$SQ_{II} / (K-1) / SQ_{res} / [n - (J + K - 1)]$	$F_{K-1, n-(J+K-1)}$
<b>terzo</b> fatt.	$SQ_I + SQ_{II} / (J + K - 2) / SQ_{res} / [n - (J + K - 1)]$	$F_{J+K-2, n-(J+K-1)}$

Invece il modello per l'ANOVA con intersezione si presenta quanto segue:

$$Y_i = \mu + \alpha_2 x_{i,2} + \dots + \alpha_J x_{i,J} + \gamma_2 x_{i,J+1} + \dots + \gamma_K x_{i,J+K-1} + \delta_1 x_{i,2} x_{i,J+1} + \dots + \varepsilon_i$$

l'ipotesi nulla sarà:  $H_0: \delta_1 = \dots = \delta_{JK}$

e la statistica test utilizzata si definisce:  $F = \frac{SQ_{int} / (J-1)(K-1)}{SQ_{res} / (n-JK)} \sim F_{(J-1)(K-1), n-JK}$

### 5.3 ANCOVA

Il test *ANCOVA* considera un modello lineare in cui sono presenti sia variabili esplicative categoriale che variabili quantitative; tale test prevede l'analisi della covarianza.

Gli obiettivi dell'analisi sono valutare la differenza in media nei diversi gruppi tenendo conto della variabile quantitativa, dunque definire la relazione tra la variabile risposta e l'esplicativa quantitativa nei differenti gruppi osservati. Se si considera una variabile quantitativa e una categoriale con due o più modalità, nel modello lineare allora viene aggiunta una variabile esplicativa risultante dall'intersezione delle due appena descritte.

Il modello di regressione lineare è in grado di stabilire se la variabile categoriale influenza la media della variabile risposta e la relazione tra la variabile risposta e la covariata quantitativa; se esiste la relazione, permette di individuare il contributo della variabile categoriale sulla risposta avendo una variazione uguale comunque sia definita la variabile quantitativa oppure avendo una variazione in base alla variabile quantitativa. Il test dunque descrive, all'interno di ciascun gruppo definito in base ai livelli del fattore, come varia la variabile risposta con un cambiamento unitario del predittore quantitativo. Il test presuppone come ipotesi nulla la ininfluenza della suddivisione nei diversi livelli del fattore, ossia definisce che tale suddivisione non ha alcun effetto sulla variabile dipendente tenendo conto della variabile quantitativa; un rifiuto dell'ipotesi nulla definisce una qualche relazione, ma non dice nulla sulla direzione degli effetti inclusi nel modello e un termine non significativo in riferimento all'effetto dell'interazione significa che i dati non supportano l'ipotesi che la variabile risposta sia aumentata in modo differente in ciascun livello del fattore. Per il test devono essere soddisfatti gli stessi requisiti richiesti dal test *ANOVA* a due fattori e in aggiunta richiede che sia verificata l'ipotesi di omoschedasticità tra i gruppi definiti in base alle modalità dalla variabile categoriale del modello. In R il test *ANCOVA* è uguale al test *ANOVA* a due fattori (*anova()*), il software R è in grado di differenziarli osservando che entrambi contengono una variabile esplicativa categoriale ma l'*ANOVA* a due fattori presenta una seconda variabile categoriale con una o più modalità mentre nell'*ANCOVA* l'esplicativa categoriale è affiancata da una variabile indipendente quantitativa.

Se consideriamo il modello di regressione includente una variabile quantitativa ( $x_{i2}$ ) e una variabile categoriale ( $x_{i3}$ ) avanti due sole modalità, allora il modello sarà:

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i2} x_{i3} + \varepsilon_i$$

Il test valuta quindi l'assenza di effetto della variabile categoriale sulla variabile risposta tenendo conto della variabile quantitativa, dunque l'assenza d'interazione che può essere espressa con l'ipotesi nulla:

$$H_0: \beta_3 = \beta_4 = 0$$

il test confronta la varianza residua del modello sotto  $H_0$  con quella del modello completo, è dunque un rapporto di varianze che si riferiscono a quanta variabilità nei dati viene spiegata quando nei dati si includono tutti i termini, i quali rispettivamente trattano l'effetto della variabile quantitativa e della variabile categoriale sulla risposta e l'effetto dell'interazione tra le due variabili citate sulla variabile risposta.

Segue la statistica test:

$$F = \frac{(\tilde{\sigma}^2 - \hat{\sigma}^2)/2}{\hat{\sigma}^2/(n-4)} \sim F_{2,n-4}$$

che include  $\tilde{\sigma}^2$  e  $\hat{\sigma}^2$  ossia rispettivamente la stima della varianza del modello ridotto e quella del modello completo.

Nell'analisi *height*, *diameter1*, *diameter2*, *number\_leaves*, *length\_leaves* e *number\_flowere* sono le variabili risposta analizzate separatamente e nei vari modelli si includono la variabile quantitativa *week* ( $x_{i2}$ ), la variabile categoriale *light* che definiscono la condizione di luce a cui è soggetto il campo, nello specifico incorpora le quattro modalità: FL, SH30, SH60, HSH60. Le modalità sono rappresentate da variabili indicatrici (dummy) e il livello FL è considerato come riferimento per l'analisi ( $x_{i1}$ )

$$x_{i1} = \text{condizione FL} = \begin{cases} 1 & \text{se FL} \\ 0 & \text{altrimenti} \end{cases}$$

$$x_{i3} = \text{condizione SH60} = \begin{cases} 1 & \text{se SH60} \\ 0 & \text{altrimenti} \end{cases}$$

$$x_{i2} = \text{condizione SH30} = \begin{cases} 1 & \text{se SH30} \\ 0 & \text{altrimenti} \end{cases}$$

$$x_{i4} = \text{condizione HSH60} = \begin{cases} 1 & \text{se HSH60} \\ 0 & \text{altrimenti} \end{cases}$$

le variabili  $x_{i2}$ ,  $x_{i3}$  e  $x_{i4}$  definiscono rispettivamente le modalità SH30, SH60 e HSH60. L'interazione fra la variabile quantitativa e la categoriale è dunque scomponibile in  $x_{i6}$ ,  $x_{i7}$  e  $x_{i8}$ .

Il modello è:

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \varepsilon_i$$

e l'ipotesi nulla è:  $H_0: \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$ .

#### 5.4 Test *ANCOVA* per ciascuna variabile

Nel paragrafo che segue si svolgeranno i test *ANCOVA* per ciascuna variabile risposta, nello specifico in riferimento a *height*, *diameter1*, *diameter2*, *number\_leaves*, *length\_leaves* e *number\_flowers*.

Si inizia con una prima analisi che sfrutta i boxplot relativi alla variabile dipendente nei differenti campi di osservazione, distinti per la condizione di luce; graficamente si cerca di cogliere le differenze tra le condizioni di luce prescindendo dal tempo, ossia trascurando lo scorrere delle settimane di osservazione. Nella realtà potrebbero non esistere delle differenze in un determinato istante temporale oppure potrebbero esistere, magari determinate in una direzione in un istante e in una direzione opposta in un secondo istante temporale, per cui globalmente tali differenze si annullerebbero; si attua quindi un'analisi più approfondita servendosi dei modelli di regressione polinomiale stimati nel capitolo precedente, ciascuno con il grado del polinomio considerato più opportuno. L'obiettivo è stabilire se esiste una qualche differenza significativa tra le diverse condizioni di luce tenendo conto dell'andamento del tempo specifico di ciascuna condizione, ossia si vuole verificare se c'è un cambiamento in media della risposta nelle diverse condizioni tenendo in considerazione il fatto che i valori delle variabili risposta aumentano con lo scorrere delle settimane.

Il procedimento consiste nell'applicazione del test *ANCOVA*, il quale confronta il modello ridotto che considera esclusivamente la variabile *week* e il modello che include anche la categoriale *light* e l'interazione fra le due, per ciascuna variabile dipendente. Il test pone come ipotesi nulla la nullità dei coefficienti che interessano le diverse condizioni di luce, esclusa quella posta come riferimento, e il coefficiente che rappresenta l'interazione.

### 5.4.1 Altezza

Trattando l'altezza delle piantine (*height*) come variabile risposta, in primo luogo si ripotano i boxplot per un'analisi marginale per ciascuna condizione di luce:

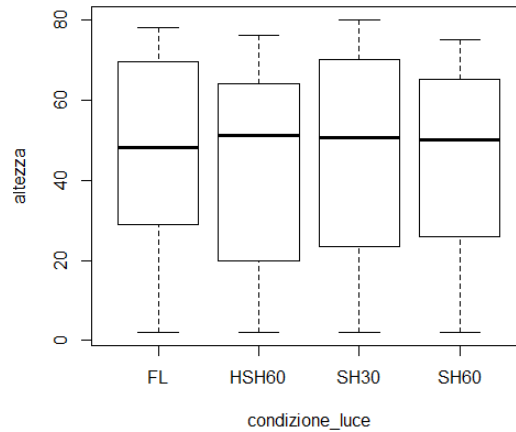


Figura 1c: Boxplot per i fattori *condizione\_luce* (*light*) rispetto alla variabile risposta *height*

Graficamente appare che le condizioni di luce influenzate da agrivoltaico sono leggermente più favorevoli per la crescita delle piantine, infatti le medie dei diversi gruppi sono leggermente superiori rispetto a quella della condizione di luce piena (FL).

Per il test si sfrutta il modello di regressione polinomiale di grado 2 stimato in precedenza, pertanto si confronta il modello ridotto con il modello completo.

Analysis of Variance Table

```
Model 1: height ~ poly(week, 2)
Model 2: height ~ poly(week, 2) * light
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     381 10721
2     372 10034   9    686.44 2.8276 0.003147 **
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Il test presenta un p-value pari a 0.003147.

Figura 2c: Output del test ANCOVA che confronta il modello ridotto (Model 1) e il modello completo (Model 2)

Il test riporta un p\_value pari a 0.003147, significativo a un livello di significatività dello 0.01, pertanto si rigetta l'ipotesi di un aumento dell'altezza uguale nei diversi campi che differiscono fra loro per la condizione di luce a cui sono soggetti con un margine errore del 1%.



### 5.4.2 Diametro Principale

Seguono i boxplot riferiti alla variabile risposta *diameter1* che rappresenta il diametro principale delle piantine:

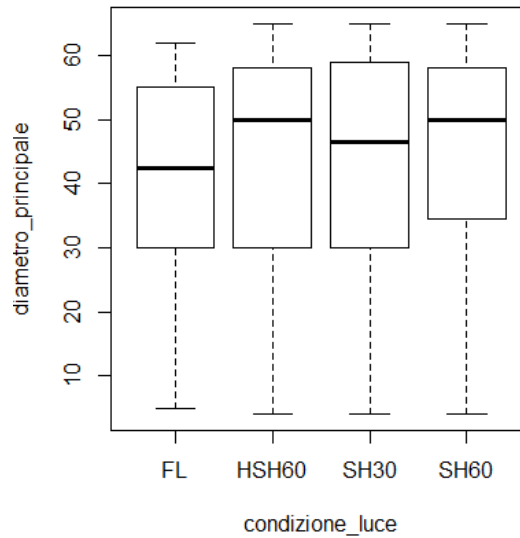


Figura 3c: Boxplot per i fattori *condizione\_luce* (light) rispetto alla variabile risposta *diameter1*

Anche per il diametro principale le condizioni artificiali favoriscono uno sviluppo maggiore, mediamente i valori assunti si accostano a quelli rilevati nelle piantine situate nei campi FL; osservando unicamente quest'ultime condizioni, SH30 appare la condizione più sfavorevole allo sviluppo del diametro principale. Graficamente si conclude che vi è una differenza nell'evoluzione media del diametro principale delle piantine e che un campo maggiormente esposto alla luce solare sembrerebbe più sfavorevole per il diametro principale.

Il modello polinomiale di secondo grado viene utilizzato nel test:

Analysis of Variance Table

```
Model 1: diameter1 ~ poly(week, 2)
Model 2: diameter1 ~ poly(week, 2) * light
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     381 14912
2     372 14124  9    787.97  2.3059 0.01569 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Il test presenta un p-value pari a 0.01569.

Figura 4c: Output del test ANCOVA che confronta il modello ridotto (Model 1) e il modello completo (Model 2)

Un p\_value pari a 0.01569 risulta superiore dello 0.01 ma inferiore allo 0.05, dunque per l'analisi si preferisce non rifiutare l'ipotesi di uno sviluppo uguale nelle diverse condizioni di luce.

### 5.4.3 Diametro Secondario

La variabile risposta analizzata riguarda il diametro secondario delle piantine (*diameter2*), i relativi boxplot seguono:

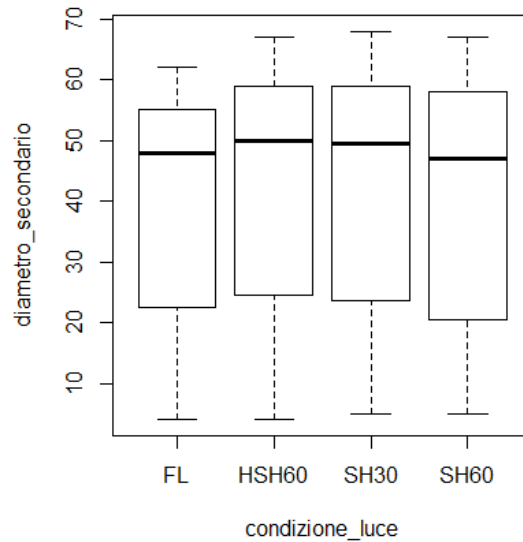


Figura 5c: Boxplot per i fattori *condizione\_luce* (light) rispetto alla variabile risposta *diameter2*

Per il diametro secondario invece la condizione di luce SH60 rivela mediamente un'evoluzione inferiore rispetto a quella del diametro secondario delle piantine allocate nei diversi campi di coltivazione. La differenza comunque appare minima anche confrontando le restanti condizioni di luce; HSH60 e SH30 sembrerebbero assumere mediamente valori leggermente superiori. Ai dati è stato adattato un modello polinomiale di grado 2, con il quale si svolge il test:

Analysis of Variance Table

```
Model 1: diameter2 ~ poly(week, 2)
Model 2: diameter2 ~ poly(week, 2) * light
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     381 15658
2     372 15054   9    603.39 1.6567 0.09786 .
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Il test presenta un p-value pari a 0.09786.

Figura 6c: Output del test ANCOVA che confronta il modello ridotto (Model 1) e il modello completo (Model 2)

Il test per il diametro secondario giunge a un p\_value pari a 0.09786 ed essendo molto elevato anche a un livello di significatività dello 0.05, non vi è evidenza empirica contro l'ipotesi nulla di uno sviluppo simultaneo dei diametri delle piantine caratterizzate dalle diverse condizioni di luce.

#### 5.4.4 Numero di foglie

La variabile risposta *number\_leaves* è una variabile di conteggio in quanto rappresenta il numero di foglie; è una risposta non normale che appartiene alla classe dei GLM.

Il boxplot riporta le scatole con baffi le quali fanno riferimento alle condizioni di luce dei campi in cui si trovano le piantine interessate; l'analisi marginale rivela che nei campi più illuminati da luce solare (FL e SH30), una piantina possiede un numero mediamente più elevato più foglie.

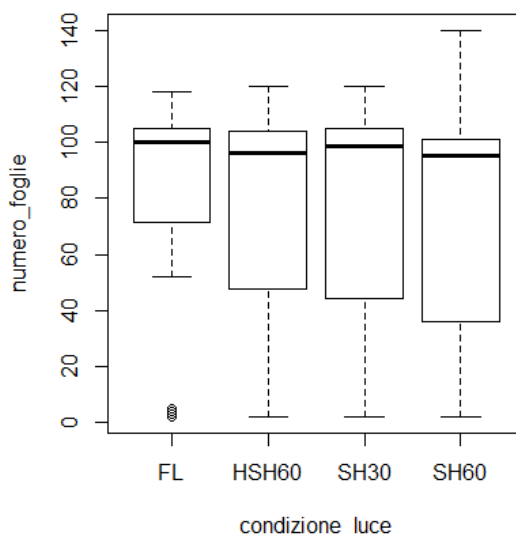


Figura 7c: Boxplot per i fattori *condizione\_luce* (light) rispetto alla variabile risposta *number\_leaves*

La differenza fra lo sviluppo di foglie nelle diverse piantine permane non sostanziale, ma lieve; per affermarlo con certezza si introduce la variabile *week* che rappresenta il tempo e si utilizza il modello di regressione con un grado del polinomio pari a 3 per svolgere il test, il quale confronta il modello ridotto con quello completo mediante il test "Chisq".

Analysis of Deviance Table

```
Model 1: number_leaves ~ poly(week, 3)
Model 2: number_leaves ~ poly(week, 3) * light
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      380    748.45
2      368    632.56 12   115.89 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Il p\_value risulta inferiore a 2.2e-16.

Figura 8c: Output del test ANCOVA che confronta il modello ridotto (Model 1) e il modello completo (Model 2)

Un p-value inferiore a 2.2e-16 appare troppo piccolo per non rifiutare l'ipotesi nulla del test, dunque si dichiara che la condizione di luce che influenza i campi coltivata ha effetto sulla variazione del numero di foglie delle piantine, cioè la suddivisione creata grazie alle diverse condizioni di luce è rilevante sulla variabile risposta (*number\_leaves*).

### 5.4.5 Lunghezza delle foglie

La seconda caratteristica delle foglie che riguarda la loro lunghezza (*length\_leaves*) viene ora trattata come variabile risposta.

Segue il boxplot per la variabile risposta *length\_leaves*:

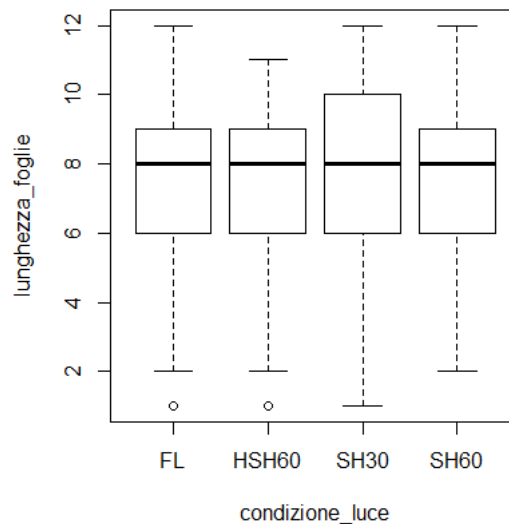


Figura 9c: Boxplot per i fattori *condizione\_luce* (*light*) rispetto alla variabile risposta *length\_leaves*

Dai boxplot emerge una lunghezza media uguale in ciascuna condizione di luce visionata. Si considera ora il modello di regressione polinomiale di grado 3 che include anche l'interazione fra la variabile esplicativa *week* e l'esplicativa *light*.

Applicando il test in riferimento alla variabile risposta *length\_leaves*, esso risulta:

Analysis of Variance Table

```
Model 1: length_leaves ~ poly(week, 3)
Model 2: length_leaves ~ poly(week, 3) * light
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1     380 726.84
2     368 641.62 12     85.219 4.0731 5.67e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Il test presenta un p-value pari a 5.67e-06.

Figura 10c: Output del test ANCOVA che confronta il modello ridotto (Model 1) e il modello completo (Model 2)

L'analisi della covarianza conduce a un p\_value largamente inferiore a 0.001, si può pertanto rifiutare l'ipotesi nulla fornita dal test e dichiarare che la variazione della lunghezza delle piantine cambia in base al campo in cui è stata allocata la piantina, ossia che esiste una diversa evoluzione della caratteristica delle foglie all'interno dei vari gruppi formatosi dalla condizione di luce del campo di appartenenza.

### 5.4.6 Numero di fiori

Il discorso discusso per la variabile risposta *number\_leaves* vale pure per la seconda variabile discreta *number\_flowers*, che indica il numero di fiori di una piantina.

L'analisi marginale del fattore *light* sulla variabile risposta *number\_flowers* mediante il boxplot rivela che la condizione di luce HSH60 è l'unica in grado di permettere alle piantine uno sviluppo mediamente significativo del numero di fiori.

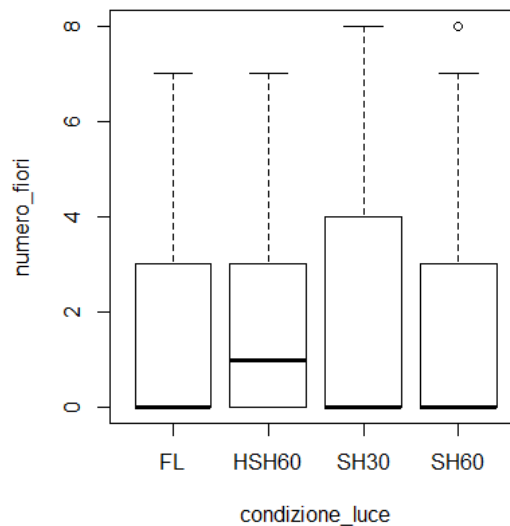


Figura 11c: Boxplot per i fattori *condizione\_luce* (*light*) rispetto alla variabile risposta *number\_flowers*

Avendo stimato il modello di regressione di grado 2 che considera le variabili esplicative quantitativa (*week*) e categoriale (*light*) e l'interazione tra le due, si svolge il test:

Analysis of Deviance Table

```
Model 1: number_flowers ~ poly(week, 2)
Model 2: number_flowers ~ poly(week, 2) * light
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         381     447.56
2         372     413.86  9   33.698 0.0001009 ***
```

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Il *p\_value* è pari a 0.0001009.

Figura 12c: Output del test ANCOVA che confronta il modello ridotto (Model 1) e il modello completo (Model 2)

Il test rigetta un *p\_value* pari a 0.0001009, il quale conduce a un rifiuto dell'ipotesi nulla, pertanto si conclude che il numero di fiori che si genera nelle piantine è differente e varia in base al campo in cui è allocata la piantina, per cui in base alla condizione di luce a cui è soggetta.

## COMMENTO

Si premette che vi è un'interazione fra la variabile *week* e le variabili risposta *height*, *diameter1*, *diameter2*, *number\_leaves*, *length\_leaves* e *number\_flowers* distinte, ossia esiste una relazione tra la variabile trattata che rappresenta una caratteristica delle piantine e lo scorrere delle settimane, conclusione che conferma l'intuizione secondo cui con il passare delle settimane la piantina cresce e aumenta lo sviluppo di foglie e fiori annesse ad essa.

L'analisi sulla covarianza valuta se esiste un'effettiva differenza della crescita e dello sviluppo medio delle piantine nei diversi gruppi formati in base alla condizione di luce a cui è soggetto il campo in cui la piantina è allocata, tenendo in considerazione il fatto che vi è una dipendenza tra la variabile *week* e la variabile risposta, qualunque essa sia. Il test confronta il modello ridotto, il quale valuta solamente l'esplicativa *week*, e il modello completo, il quale considera come esplicative le variabili *week* e *light* e l'interazione fra le due.

Per l'altezza delle piantine e per la variazione del numero di fiori e di foglie, oltre che per la lunghezza dell'ultimo attributo menzionato delle piantine, si evolvono mediamente in modo diverso in ciascun campo sottoposto a osservazione, ossia la conclusione a cui si giunge è che la suddivisione in gruppi, nello specifico nei vari campi differenti fra loro per la condizione di luce, ha effetto sulle distinte variabili risposta, dunque che è utile se si vuole favorire la crescita o lo sviluppo di determinate caratteristiche delle piantine.

L'evoluzione dei due diametri invece non varia in base alla condizione di luce a cui è soggetta la piantina in questione, dunque un'eventuale suddivisione in campi distinti non altera lo sviluppo dei diametri.

## 6. I FATTORI METEO

Tenendo presente che gli agenti del meteo alternano l'umidità e altri fattori rilevanti, l'obiettivo è stabilire se il meteo influenzano la crescita delle piantine e delle sue componenti sotto le quattro condizioni di luce che determinano i diversi campi di osservazione, dunque trovare la posizione dell'impianti agrivoltaico che ottimizza lo sviluppo delle piantine, ossia la condizione di luce che più si avvicina alla condizione di luce piena (FL). Il procedimento consiste nel verificare se le condizioni meteo in una settimana ha effetto sulla crescita delle piantine nella settimana successiva.

### 6.1 Analisi delle componenti principali (PCA)

L'analisi delle componenti principale (*Principal Component Analysis*) è una tecnica statistica multivariata applicata ai modelli per ridurre la loro dimensione, ossia per sintetizzare l'informazione fornita dalle variabili in un numero ridotto di componenti principali minimizzando la perdita d'informazione; viene applicato alle variabili originali del dataset per generare nuove variabili rilevanti incorrelate fra loro, quindi produce un modello più parsimonioso che rispecchia il principio di Rasoio di Occam, secondo il quale il modello migliore è quello che pur avendo meno variabili esplicative è in grado di spiegare bene i dati osservati.

L'auspicio dell'analisi delle componenti principali è che le varianze campionarie delle prime componenti principali siano tanto grandi da spiegare bene il fenomeno e quindi che le rimanenti componenti abbiano varianza tanto piccola da considerarle trascurabili; una variabile che riporta una scarsa variabilità la si può approssimare a una costante e dunque eliminare dal modello.

La PCA prevede l'utilizzo del vettore reale  $a = (a_1, \dots, a_p)^T$  di dimensione  $(p \times 1)$  per generare combinazioni lineari  $y_i$  del vettore contenente le  $p$  variabili originali di dimensione  $(p \times 1)$ .

$$\begin{aligned} y_1 &= a_1^T x = a_{11}x_1 + \dots + a_{1p}x_p \\ &\cdot \\ &\cdot \\ y_p &= a_p^T x = a_{p1}x_1 + \dots + a_{pp}x_p \end{aligned}$$

Ciascun  $y_i$  è una variabile casuale pertanto:

- $E[y_i] = a_i^T \mu$
- $V(y_i) = a_i^T \Sigma a_i$
- $cov(y_i, y_j) = a_i^T \Sigma a_j$

$y_i$  identifica la  $i$ -esima combinazione lineare per un generico vettore di osservazioni le cui componenti rappresentano le variabili esplicative per un'unica unità statistica; dunque considerando l'intero dataset  $X$ , il vettore delle combinazioni per ciascuna componente del vettore reale  $a$  risulta  $Y_i = Xa_i$  e la matrice dell'insieme delle combinazioni è  $Y = XA$ .

Il vettore  $a = (a_1, \dots, a_p)^T$  è specificato in modo da rispettare le seguenti condizioni:

- ciascuna componente lineare deve avere varianza massima

$$\max V(y_i) = a_i^T \Sigma a_i = \lambda_i$$

- le componenti lineari devono essere incorrelate fra loro

$$cov(y_i, y_j) = a_i^T \Sigma a_j = 0$$

sotto il vincolo di determinantalità che imposta  $a_i^T a_i = 1$  in quanto la varianza è illimitata e sotto il vincolo di incorrelazione della combinazione lineare osservata con le precedenti; se i vincoli vengono entrambi rispettati allora le combinazioni lineari sono chiamate componenti principali (CP).

### *Decomposizione spettrale DS*

La decomposizione spettrale (*eigendecomposition*) è una scomposizione di una generica matrice simmetrica in autovalori e autovettori, nello specifico  $A$  una matrice ( $p \times p$ ), può essere definita:

$$A = \sum_{i=1}^p \lambda_i \gamma_i \gamma_i^T = \Gamma \Lambda \Gamma^T$$

con  $\Gamma$  matrice degli autovettori e  $\Lambda$  matrice degli autovalori.



Gli autovettori  $\gamma_i$  indicano le direzioni ortogonali di massima variabilità di ciascuna componente principale e son utili per la generazione delle stesse, dunque sostituiscono le componenti del vettore  $a$ ; i singoli autovalori  $\lambda_i$  disposti in ordine decrescente rappresentano invece la varianza massima delle singole componenti principali.

#### *Decomposizione a valori singolari (svd)*

La decomposizione a valori singolari (*Singular Value Decomposition*) non richiede che la matrice sia simmetrica, dunque data  $A$  una matrice qualsiasi ( $n \times p$ ), permette la definizione della stessa come:

$$A = \sum_{i=1}^p d_i u_i v_i^T = U D V^T$$

La decomposizione si svolge attuando una DS alla matrice simmetrica  $AA^T$  da cui si ricava  $U$  matrice degli autovettori,  $D$  la matrice diagonale i cui elementi sono la radice degli autovalori non nulli ottenuti dalla DS e  $V$  la matrice composta da  $v_i = d_i^{-1} A^T u_i$  con  $i = 1, \dots, \min(n, p)$ .

#### *Standardizzazione*

Una proprietà importante dell'analisi delle componenti principali (PCA) è che non è invariante rispetto alla standardizzazione, il che vuol dire che se si considera una trasformazione  $Z$  di posizione e scala per il vettore  $X$  di partenza ottenuta come:

$$Z = \Delta^{-1/2} (X - \mu) \quad \text{con } \Delta^{-1/2} = \text{diag}(\text{diag}(S))^{-1/2}$$

Occorre applicare la decomposizione spettrale alla matrice di correlazione  $R$  anziché alla matrice di varianze e covarianze  $S$ . La standardizzazione di  $X$  consente alle variabili contenute al suo interno di avere la medesima variabilità e si applica quando le prime CP hanno valori  $\lambda_i$  troppo elevati, ossia assorbono più variabilità delle restanti CP, ma il suo utilizzo viene richiamato anche quando le unità di misura sono diverse oppure quando appare troppa differenza tra i vari criteri citati in precedenza.

### Procedimento

Dato un dataset  $X$  contenute le variabili esplicative quantitative originali distribuito secondo una distribuzione con una certa media e una certa varianza e dato il campione di osservazione, il procedimento manuale prevede di determinare alcune statistiche descrittive utili ai fini dell'analisi, in particolare:

- media campionaria per stimare la media della distribuzione di  $X$

$$\bar{x} = \frac{1}{n} X^T \mathbf{1}$$

- varianza non corretta per stimare la varianza della distribuzione di  $X$

$$S = \frac{1}{n} X^T H X$$

con  $H = I - \frac{1}{n} \mathbf{1}^T \mathbf{1}$  matrice di centramento

- correlazione

$$R = D^{-1/2} S D^{-1/2}$$

con  $D = \text{diag}(S)$

Successivamente si applica la decomposizione spettrale DS alla matrice di varianza e covarianze  $S$ , essendo una matrice simmetrica, ottenendo così gli autovettori e gli autovalori della matrice necessari per la tecnica, dunque calcolano le componenti principali CP trasformando i vettori della matrice dei dati  $X$  servendosi della DS ricavata e le si centrano premoltiplicandole per la matrice di centramento  $H$ .

Il numero di componenti principale da considerare si determina attraverso vari metodi, si riportano in seguito alcuni di essi:

- Data la proporzione di varianza totale spiegata

$$V(Y_i) = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_p}$$

ossia quanto la componente principale informa riguardo alla qualità della PCA applicata ai dati, il metodo consiste nell'includere tante variabili quante insieme sono necessarie per spiegare la varianza totale  $V(Y_i)$ ; occorre tener presente che  $Y_1$  è più importante di  $Y_2$  in quanto  $\lambda_1 > \lambda_2$ , con  $\lambda_1$  varianza di  $Y_1$  e  $\lambda_2$  varianza di  $Y_2$ .

- Un secondo metodo sfrutta il grafico screenplot che riporta la varianza, ossia gli autovalori  $\lambda_i$  corrispondenti a ciascuna CP, dunque attraverso l'entità di  $\lambda_i$  si individua il punto di gomito in cui l'andamento della curva cambia di direzione, quindi si considerano  $(i - 1)$  CP oppure si considerano tante CP quante sono quelle caratterizzate da una varianza  $\lambda_i > 1$ .
- L'individuazione del numero di componenti principali rilevanti si può cogliere anche dall'interpretazione di esse, ossia stabilendo quali sono legate alle variabili osservando la correlazione tra ogni CP e ogni variabile casuale oppure osservando la matrice gamma  $\Gamma$ ; valori elevati indicano che la variabile casuale in questione determina la corrispettiva CP.

L'auspicio dell'analisi delle componenti principali è che le varianze campionarie delle prime componenti principali siano tanto grandi da spiegare bene il fenomeno e quindi che le rimanenti componenti abbiano varianza tanto piccola da considerarle trascurabili; una variabile che riporta una scarsa variabilità la si può approssimare a una costante e dunque eliminare dal modello.

## SOFTWARE R

In R Studio la funzione per applicare l'analisi delle componenti principali (PCA) al dataset d'interesse è ***prcomp()*** la quale svolge la decomposizione a valori singolari (*svd*) sulla matrice di varianze e covarianze; la funzione restituisce la deviazione standard della componenti principali, in altri termine la radice degli autovalori (*.\$sdev*), la matrice di rotazione ossia la matrice gamma  $\Gamma$  contenente gli autovettori della matrice di varianze e covarianze (*.\$rotation*), valori maggiori stanno a indicare che la componente principale è determinata dalle variabili in corrispondenza di quei valori. La funzione inoltre restituisce il vettore delle medie (*.\$center*), la deviazione standard delle variabili (*.\$scale*) e la matrice delle componenti principali centrate (*.\$x*).

Osservando la matrice degli autovalori, se essi risultano diversi tra loro ossia se vi sono variabili che hanno variabilità troppo elevata rispetto a quella delle altre variabili, occorre standardizzare i dati, generando così variabili con la medesima variabilità. La tecnica non è invariante rispetto a standardizzazione, si utilizza il comando ***prcomp()*** e al suo interno si specifica l'argomento *scale="TRUE"*; la funzione standardizza i dati prima dell'applicazione della PCA per poi svolgere la decomposizione spettrale alla matrice di correlazione. Mediante il comando *summary()* gli elementi sono i medesimi della funzione *prcomp()*.

L'analisi grafica ausiliaria richiede l'utilizzo del grafico *screenplot* che presenta la varianza per ciascuna componente principale; è utile ai fine dell'individualizzazione del numero di componenti in quanto privilegia le componenti con varianza maggiore di 1, oppure le prime  $(K - 1)$  componenti dove  $K$  indica la componente in prossimità della quale si presenta un punto di gomito, ossia un punto che determina un cambio di direzione della curva.

Un secondo grafico è *biplot()* il quale figura le unità statistiche ossia la matrice delle componenti principali e le variabili d'interesse ossia gli autovettori.

## 6.2 Le variabili del dataset

Il dataset Meteo comprende valutazioni orarie per ciascun giorno dell'intervallo temporale rilevato, le cui variabili descrivono i fattori meteo; per ciascuno si riportano più variabili riferite al medesimo fattore, essi differiscono fra loro per il metodo di rilevazione, si sceglie così di trascurarne alcune e per ogni fattore si considera solamente la variabile che rappresenta la media.

Nello specifico si tiene conto delle variabili:

- **temp\_aria\_media:** la temperatura dell'aria media
- **umid\_aria\_media:** l'umidità dell'aria media
- **p\_rugiada\_media:** il punto medio di rugiada
- **umid\_suolo\_media:** l'umidità media del suolo
- **temp\_suolo\_media:** la temperatura media del suolo
- **rad\_solare2\_media:** la radiazione solare media

L'obiettivo è capire se il meteo influenza la crescita della coltura, dunque occorre studiare i fattori meteo rilevati nei giorni pregressi alla rilevazione della caratteristica della piantina. Ciascuna condizione di luce è caratterizzata da un proprio dataset indipendente dagli altri; in primo luogo per ogni dataset si selezionano le variabili menzionate in precedenza ottenendo così un dataset ridotto con rilevazioni orarie, dunque si effettua una media giornaliera per ogni variabile d'interesse riducendo in tal modo ciascun dataset a 96 unità statistiche.

### **6.3 Applicazione della PCA per ciascuna condizione di luce**

Viene applicata l'analisi delle componenti principali a ciascun dataset ognuno caratterizzato dai fattori meteo rilevati nei diversi campi suddivisi per condizione di luce a cui sono soggetti. In riferimento a ciascuna condizione di luce si seleziona il numero di componenti principali in grado di spiegare opportunamente i dati mediante il grafico *screenplot* e l'osservazione della varianza di ogni componente principale.

In riferimento a una determinata variabile risposta, per ciascuna condizione di luce visionata si stima dunque un modello che include un polinomio con il grado considerato più adeguato al contesto e le componenti principali selezionate; dalle analisi la prima componente principale consegue in grado di spiegare i dati e buona parte della variabilità totale in ogni condizione di luce.

Si commentano i relativi output dei modelli risultanti per verificare se i fattori meteo dei giorni precedenti alle rilevazioni influenzano in qualche modo la crescita e lo sviluppo delle peculiarità delle piantine nei differenti campi di coltivazione.

### 6.3.1 Altezza

Per ciascuna condizione di luce visionata, si riportano gli output dei modelli polinomiali di secondo ordine per la variabile risposta che riguarda l'altezza delle piantine (*height*) aggiunta della variabile esplicativa che rappresenta la prima componente principale selezionata.

#### FL

```
Call:
lm(formula = height ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-12.3758 -3.0933 -0.4497  3.7446  9.7138

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -5.551    12.208  -0.455   0.65
poly(week, 2)1 163.094    17.399   9.374 4.75e-15 ***
poly(week, 2)2 -51.839     4.869  -10.647 < 2e-16 ***
PC1            -21.740     5.208  -4.174 6.78e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.785 on 92 degrees of freedom
Multiple R-squared:  0.9643,    Adjusted R-squared:  0.9632
F-statistic: 829.2 on 3 and 92 DF,  p-value: < 2.2e-16
```

#### SH30

```
Call:
lm(formula = height ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-16.5126 -2.9772  0.0228  3.0702 12.4874

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -11.661    15.131  -0.771 0.442887
poly(week, 2)1 172.714    18.841   9.167 1.29e-14 ***
poly(week, 2)2 -40.958     5.712  -7.170 1.84e-10 ***
PC1           -24.335     6.443  -3.777 0.000282 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.363 on 92 degrees of freedom
Multiple R-squared:  0.9583,    Adjusted R-squared:  0.9569
F-statistic: 704.7 on 3 and 92 DF,  p-value: < 2.2e-16
```

#### SH60

```
Call:
lm(formula = height ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
 -9.8646 -1.7051 -0.0858  2.1574 13.1354

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   37.487     7.440   5.038 2.33e-06 ***
poly(week, 2)1 218.179     7.436  29.341 < 2e-16 ***
poly(week, 2)2 -53.629     3.992  -13.434 < 2e-16 ***
PC1            -2.921     3.184  -0.917   0.361
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.952 on 92 degrees of freedom
Multiple R-squared:  0.9736,    Adjusted R-squared:  0.9728
F-statistic: 1133 on 3 and 92 DF,  p-value: < 2.2e-16
```

#### HSH60

```
Call:
lm(formula = height ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-21.1730 -3.4926 -0.1605  2.9029 15.3933

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   44.8603    15.2252  2.946 0.00407 **
poly(week, 2)1 229.3182    16.3241 14.048 < 2e-16 ***
poly(week, 2)2 -47.0228     6.3928  -7.356 7.7e-11 ***
PC1             0.9431     6.4796  0.146 0.88459
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.731 on 92 degrees of freedom
Multiple R-squared:  0.9468,    Adjusted R-squared:  0.9451
F-statistic: 546.2 on 3 and 92 DF,  p-value: < 2.2e-16
```

Figura 1d: Output dei modelli stimati per la variabile *height* in ciascuna condizione di luce

Osservando i coefficienti che interessano la componente principale (PC1), solo quelli nelle condizioni di FL e SH30 risultano significativi, ciò significa che nelle condizioni di luce restanti il coefficiente in questione può essere omesso.

Si può affermare di conseguenza che i fattori meteo dei giorni precedenti alla rilevazione influenzano la crescita delle piantine solamente nei campi FL e SH30, esposti maggiormente alla luce solare e di conseguenza agli agenti atmosferici in generale.

### 6.3.2 Diametro Principale

Il diametro principale (*diameter1*) è stato adattato un modello polinomiale di ordine 2, al quale viene aggiunta la prima componente principale; seguono gli output distinti per condizione.

#### FL

```
Call:
lm(formula = diameter1 ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-9.7428 -3.1059  0.0727  3.5892 12.7959

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   33.559    13.002   2.581  0.0114 *
poly(week, 2)1 154.619    18.532   8.343 6.92e-13 ***
poly(week, 2)2 -46.789     5.186  -9.022 2.60e-14 ***
PC1           -2.581     5.547  -0.465  0.6428
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.097 on 92 degrees of freedom
Multiple R-squared:  0.9231,    Adjusted R-squared:  0.9206
F-statistic: 368.2 on 3 and 92 DF,  p-value: < 2.2e-16
```

#### SH30

```
Call:
lm(formula = diameter1 ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-13.1777 -3.8429 -0.4473  2.7265 26.6442

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   52.673    20.190   2.609  0.0106 *
poly(week, 2)1 177.458    25.140   7.059 3.10e-10 ***
poly(week, 2)2 -54.125     7.622  -7.101 2.54e-10 ***
PC1            4.668     8.598   0.543  0.5885
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.155 on 92 degrees of freedom
Multiple R-squared:  0.8636,    Adjusted R-squared:  0.8591
F-statistic: 194.1 on 3 and 92 DF,  p-value: < 2.2e-16
```

#### SH60

```
Call:
lm(formula = diameter1 ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-20.6032 -3.9273 -0.0385  3.4080 17.0956

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   37.446    11.830   3.165  0.0021 **
poly(week, 2)1 156.541    11.822  13.241 <2e-16 ***
poly(week, 2)2 -68.310     6.347 -10.762 <2e-16 ***
PC1           -2.117     5.063  -0.418  0.6768
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.284 on 92 degrees of freedom
Multiple R-squared:  0.8938,    Adjusted R-squared:  0.8903
F-statistic: 258 on 3 and 92 DF,  p-value: < 2.2e-16
```

#### HSH60

```
Call:
lm(formula = diameter1 ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-13.6259 -2.9830 -0.6838  2.2212 17.3162

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   27.562    15.999   1.723  0.0883 .
poly(week, 2)1 154.931    17.154   9.032 2.49e-14 ***
poly(week, 2)2 -68.984     6.718 -10.269 < 2e-16 ***
PC1           -5.865     6.809  -0.861  0.3913
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.022 on 92 degrees of freedom
Multiple R-squared:  0.908,    Adjusted R-squared:  0.905
F-statistic: 302.6 on 3 and 92 DF,  p-value: < 2.2e-16
```

Figura 2d: Output dei modelli stimati per la variabile *diameter1* in ciascuna condizione di luce

Per le condizioni di luce analizzate la prima componente principale non apporta miglioramenti ai singoli modelli ridotti, a indice che il diametro principale delle piantine situate in ogni campo osservato non è sensibile il meteo dei giorni pregressi alla rilevazione.



### 6.3.3 Diametro Secondario

Anche per il diametro secondario (*diameter2*) si riporta gli output dei modelli che includono un polinomio di grado 2 e la prima componente principale per ciascuna condizione di luce.

#### FL

```
Call:
lm(formula = diameter2 ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-10.1933  -2.6033   0.3396   2.4031   9.1671

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    60.497    10.732   5.637 1.88e-07 ***
poly(week, 2)1  198.180    15.296  12.956 < 2e-16 ***
poly(week, 2)2  -64.181     4.280  -14.994 < 2e-16 ***
PC1              9.028     4.579   1.972  0.0517 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.207 on 92 degrees of freedom
Multiple R-squared:  0.953,    Adjusted R-squared:  0.9515
F-statistic: 621.8 on 3 and 92 DF,  p-value: < 2.2e-16
```

#### SH60

```
Call:
lm(formula = diameter2 ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-19.0982  -3.2503  -0.0527   2.9473  21.9018

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    18.368    12.814   1.433  0.155
poly(week, 2)1  159.119    12.806  12.425 < 2e-16 ***
poly(week, 2)2  -52.517     6.875  -7.638 2.02e-11 ***
PC1              -9.062     5.484  -1.652  0.102
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.807 on 92 degrees of freedom
Multiple R-squared:  0.8898,    Adjusted R-squared:  0.8862
F-statistic: 247.5 on 3 and 92 DF,  p-value: < 2.2e-16
```

#### SH30

```
Call:
lm(formula = diameter2 ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-12.4143  -3.4143  -0.2956   2.3121  21.8230

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)     30.694    17.994   1.706  0.0914 .
poly(week, 2)1  161.621    22.405   7.213 1.50e-10 ***
poly(week, 2)2  -65.497     6.793  -9.642 1.29e-15 ***
PC1              -4.809     7.662  -0.628  0.5318
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.377 on 92 degrees of freedom
Multiple R-squared:  0.9038,    Adjusted R-squared:  0.9007
F-statistic: 288.2 on 3 and 92 DF,  p-value: < 2.2e-16
```

#### HSH60

```
Call:
lm(formula = diameter2 ~ poly(week, 2) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-19.0325  -2.9722   0.4529   3.0630  21.0687

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    -11.291    19.234  -0.587  0.55862
poly(week, 2)1  122.804    20.622   5.955 4.70e-08 ***
poly(week, 2)2  -68.251     8.076  -8.451 4.12e-13 ***
PC1             -22.284     8.186  -2.722  0.00775 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.24 on 92 degrees of freedom
Multiple R-squared:  0.8776,    Adjusted R-squared:  0.8736
F-statistic: 219.8 on 3 and 92 DF,  p-value: < 2.2e-16
```

Figura 3d: Output dei modelli stimati per la variabile *diameter2* in ciascuna condizione di luce

I coefficienti della prima componente principale per le condizioni di luce SH30 e SH60 non sono significativi, per le condizioni FL e HSH60 invece risultano significativi a un livello di significatività rispettivamente uguale allo 0.1 e allo 0.01. In HSH60 il diametro secondario muta in base ai fattori del meteo, asserzione che si può trarre con un margine d'errore dell'1%; la medesima conclusione la si può dichiarare anche per il diametro delle piantine situate nel campo FL, ma il possibile errore aumenta raggiungendo il 10% e con un errore così elevato si preferisce smentire tale affermazione.

### 6.3.4 Numero di foglie

La funzione *glm* stima un modello polinomiale di ordine 3 addizionato con la prima componente principale per la variabile risposta che rappresenta il numero di foglie (*number\_leaves*); si riportano gli output dei modelli di ciascun campo oggetto di studio.

#### FL

```
Call:
glm(formula = number_leaves ~ poly(week, 3) + PC1, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9171 -0.9578 -0.1946  0.8206  2.1735

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.2220    0.6127  -3.627  0.000287 ***
poly(week, 3)1 -0.6808    0.7536  -0.903  0.366328
poly(week, 3)2 -6.7056    0.2578 -26.008 < 2e-16 ***
poly(week, 3)3  5.5221    0.3275  16.861 < 2e-16 ***
PC1          -2.6830    0.2581 -10.395 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2584.71  on 95  degrees of freedom
Residual deviance: 101.02  on 91  degrees of freedom
AIC: 671.53

Number of Fisher Scoring iterations: 4
```

#### SH60

```
Call:
glm(formula = number_leaves ~ poly(week, 3) + PC1, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3156 -0.8234 -0.2696  0.5183  4.2934

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.6475    0.3882  11.973 < 2e-16 ***
poly(week, 3)1  9.5668    0.3534  27.069 < 2e-16 ***
poly(week, 3)2 -6.2517    0.2361 -26.481 < 2e-16 ***
poly(week, 3)3  2.0434    0.2958   6.909  4.88e-12 ***
PC1           0.3095    0.1639   1.889   0.0589 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2927.72  on 95  degrees of freedom
Residual deviance: 119.34  on 91  degrees of freedom
AIC: 679.62

Number of Fisher Scoring iterations: 4
```

#### SH30

```
Call:
glm(formula = number_leaves ~ poly(week, 3) + PC1, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0967 -0.4835 -0.1863  0.6577  2.5075

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.7911    0.6763  1.170  0.242
poly(week, 3)1  5.6392    0.6710  8.404 < 2e-16 ***
poly(week, 3)2 -6.5195    0.2561 -25.455 < 2e-16 ***
poly(week, 3)3  3.8443    0.3620  10.620 < 2e-16 ***
PC1          -1.3343    0.2830  -4.715  2.41e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2852.600  on 95  degrees of freedom
Residual deviance:   78.692  on 91  degrees of freedom
AIC: 640.52

Number of Fisher Scoring iterations: 4
```

#### HSH60

```
Call:
glm(formula = number_leaves ~ poly(week, 3) + PC1, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0894 -0.5658  0.0224  0.4909  3.5135

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.5494    0.4874  1.127  0.26
poly(week, 3)1  6.2528    0.3783  16.527 < 2e-16 ***
poly(week, 3)2 -7.9459    0.3937 -20.181 < 2e-16 ***
poly(week, 3)3  3.9767    0.2878  13.817 < 2e-16 ***
PC1          -1.4226    0.2010  -7.078  1.47e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2807.42  on 95  degrees of freedom
Residual deviance: 140.31  on 91  degrees of freedom
AIC: 699.07

Number of Fisher Scoring iterations: 4
```

Figura 4d: Output dei modelli stimati per la variabile *number\_leaves* in ciascuna condizione di luce

La condizione di luce SH60 consente di stimare un modello la cui esplicativa che rappresenta la componente principale (PC1) risulta significativa a un livello pari allo 0.1, quindi si preferisce considerare il modello ridotto; la componente nei modelli non menzionati invece risultano tutti significativi.

Seguono i test *ANOVA* per ogni condizione:

### FL

```
Analysis of Deviance Table

Model 1: number_leaves ~ poly(week, 3)
Model 2: number_leaves ~ poly(week, 3) + PC1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         92    213.15
2         91    101.02  1   112.12 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### SH30

```
Analysis of Deviance Table

Model 1: number_leaves ~ poly(week, 3)
Model 2: number_leaves ~ poly(week, 3) + PC1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         92    101.788
2         91    78.692  1   23.096 1.541e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### SH60

```
Analysis of Deviance Table

Model 1: number_leaves ~ poly(week, 3)
Model 2: number_leaves ~ poly(week, 3) + PC1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         92    122.91
2         91    119.34  1    3.5598  0.05919 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### HSH60

```
Analysis of Deviance Table

Model 1: number_leaves ~ poly(week, 3)
Model 2: number_leaves ~ poly(week, 3) + PC1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         92    194.72
2         91    140.31  1   54.404 1.632e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 5d: Output del test *ANOVA* che confronta il modello ridotto (Model 1) e il modello completo (Model 2)

Dai test *ANOVA* tutti i *p*-value assumono un valore molto piccolo, quindi si conclude che il numero di foglie di ciascuna piantina allocata in campi esposti a una quantità ottimale di luce è sensibile al meteo dei giorni che precedono la rilevazione. In SH60 invece l'ombreggiamento impatta maggiormente sul terreno e la struttura dell'impianto agrivoltaico protegge maggiormente le piantine, per il seguente motivo i fattori del meteo non incidono significativamente sulla variazione del numero di foglie.

### 6.3.5 Lunghezza delle foglie

In seguito vengono presentati gli output dei modelli polinomiali cubici comprendenti anche la prima componente principale per ogni condizione di luce per la variabile risposta che identifica la lunghezza delle foglie (*length\_leaves*).

#### FL

```
Call:
lm(formula = length_leaves ~ poly(week, 3) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-2.9789 -0.9504 -0.0355  0.9488  5.1186

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -18.099     6.379  -2.837 0.005608 **
poly(week, 3)1 -15.129     8.851  -1.709 0.090820 .
poly(week, 3)2 -13.583     1.440  -9.432 3.93e-15 ***
poly(week, 3)3  11.342     2.502   4.533 1.76e-05 ***
PC1           -10.748     2.723  -3.947 0.000156 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.361 on 91 degrees of freedom
Multiple R-squared:  0.7659,    Adjusted R-squared:  0.7557
F-statistic: 74.45 on 4 and 91 DF,  p-value: < 2.2e-16
```

#### SH30

```
Call:
lm(formula = length_leaves ~ poly(week, 3) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-2.3702 -0.7818 -0.2468  0.6986  4.6298

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -16.509     4.998  -3.303 0.001369 **
poly(week, 3)1 -4.818     6.092  -0.791 0.431039
poly(week, 3)2 -4.707     1.383  -3.404 0.000988 ***
poly(week, 3)3  11.365     1.772   6.414 6.24e-09 ***
PC1           -10.311     2.129  -4.842 5.23e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.22 on 91 degrees of freedom
Multiple R-squared:  0.8384,    Adjusted R-squared:  0.8313
F-statistic: 118 on 4 and 91 DF,  p-value: < 2.2e-16
```

#### SH60

```
Call:
lm(formula = length_leaves ~ poly(week, 3) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-3.1725 -0.5018 -0.0317  0.8275  2.8275

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.667     3.087   0.540 0.590522
poly(week, 3)1 17.713     2.867   6.177 1.80e-08 ***
poly(week, 3)2 -4.463     1.199  -3.721 0.000344 ***
poly(week, 3)3  8.483     1.641   5.170 1.38e-06 ***
PC1            -2.375     1.322  -1.797 0.075706 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.177 on 91 degrees of freedom
Multiple R-squared:  0.8195,    Adjusted R-squared:  0.8116
F-statistic: 103.3 on 4 and 91 DF,  p-value: < 2.2e-16
```

#### HSH60

```
Call:
lm(formula = length_leaves ~ poly(week, 3) + PC1)

Residuals:
    Min       1Q   Median       3Q      Max
-3.2580 -0.6116  0.0801  0.7420  2.7420

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.810     3.418  -0.530 0.597593
poly(week, 3)1  8.959     3.639   2.462 0.015700 *
poly(week, 3)2 -10.963     1.368  -8.013 3.6e-12 ***
poly(week, 3)3  4.654     1.287   3.617 0.000489 ***
PC1           -3.739     1.455  -2.570 0.011779 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.211 on 91 degrees of freedom
Multiple R-squared:  0.761,    Adjusted R-squared:  0.7505
F-statistic: 72.46 on 4 and 91 DF,  p-value: < 2.2e-16
```

Figura 6d: Output dei modelli stimati per la variabile *length\_leaves* in ciascuna condizione di luce

Il coefficiente PC1, ossia quello che riguarda la componente principale è significativo nelle condizioni FL e SH30. In HSH60 e in SH60 permane significativo ma un livello di significatività rispettivamente dello 0.05 e dello 0.1. La lunghezza delle foglie dunque risente dei fattori meteo dei giorni precedenti nei campi sottoposti a maggior esposizione, in HSH60 invece, e soprattutto in SH60, la seguente affermazione la si può dichiarare con una possibilità di errore largamente maggiore, di conseguenza in entrambi la lunghezza non è influenzata dal meteo.

### 6.3.6 Numero di fiori

Si adatta ai dati un polinomio di ordine 2 includente la prima componente principale, seguono gli output dei modelli stimati che differiscono per condizione di luce a cui è sottoposto il campo.

#### FL

```
Call:
glm(formula = number_flowers ~ poly(week, 2) + PC1, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1795 -0.6973  0.0000  0.0000  2.8451

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    865.4    58462.3   0.015   0.988
poly(week, 2)1  3430.6    232521.1  0.015   0.988
poly(week, 2)2 -1209.4     81618.2 -0.015   0.988
PC1             439.0     29700.9  0.015   0.988

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 313.74 on 95 degrees of freedom
Residual deviance: 104.85 on 92 degrees of freedom
AIC: 207.51

Number of Fisher Scoring iterations: 22
```

#### SH60

```
Call:
glm(formula = number_flowers ~ poly(week, 2) + PC1, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1820 -0.9927  0.0000  0.4047  1.6058

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    226.2     190.2   1.189   0.234
poly(week, 2)1  928.5     758.2   1.225   0.221
poly(week, 2)2 -468.1     374.1  -1.251   0.211
PC1             120.5     100.7   1.196   0.232

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 285.777 on 95 degrees of freedom
Residual deviance: 76.001 on 92 degrees of freedom
AIC: 215.61

Number of Fisher Scoring iterations: 12
```

#### SH30

```
Call:
glm(formula = number_flowers ~ poly(week, 2) + PC1, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3283 -0.9016 -0.6841  0.1893  3.3511

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -7.440     2.858  -2.603  0.00924 **
poly(week, 2)1  6.545     6.091  1.074  0.28260
poly(week, 2)2 -15.577     2.212  -7.042  1.89e-12 ***
PC1            -2.903     1.253  -2.318  0.02047 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 355.85 on 95 degrees of freedom
Residual deviance: 113.17 on 92 degrees of freedom
AIC: 259.16

Number of Fisher Scoring iterations: 6
```

#### HSH60

```
Call:
glm(formula = number_flowers ~ poly(week, 2) + PC1, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7289 -0.4291 -0.1082  0.2379  2.7692

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)     3.140     3.157  0.994  0.320031
poly(week, 2)1  27.190     7.807  3.483  0.000496 ***
poly(week, 2)2 -17.339     2.437  -7.116  1.11e-12 ***
PC1              1.831     1.462  1.253  0.210284
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 253.927 on 95 degrees of freedom
Residual deviance: 78.699 on 92 degrees of freedom
AIC: 232.24

Number of Fisher Scoring iterations: 6
```

Figura 7d: Output dei modelli stimati per la variabile `number_flowers` in ciascuna condizione di luce

PC1 ossia il coefficiente che identifica la prima componente principale non è significativa in nessun modello stimato, più precisamente mostra una significatività nel modello SH30 ma a un livello di significatività dello 0.05; si opta allora per omettere la componente principale in ciascun modello riportato.

Seguono i test ANOVA per ogni condizione:

### FL

```
Analysis of Deviance Table

Model 1: number_flowers ~ poly(week, 2)
Model 2: number_flowers ~ poly(week, 2) + PC1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1          93    135.87
2          92    104.85 1    31.019 2.555e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### SH30

```
Analysis of Deviance Table

Model 1: number_flowers ~ poly(week, 2)
Model 2: number_flowers ~ poly(week, 2) + PC1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1          93    118.28
2          92    113.17 1    5.1139 0.02374 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### SH60

```
Analysis of Deviance Table

Model 1: number_flowers ~ poly(week, 2)
Model 2: number_flowers ~ poly(week, 2) + PC1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1          93    79.203
2          92    76.001 1    3.2016 0.07357 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### HSH60

```
Analysis of Deviance Table

Model 1: number_flowers ~ poly(week, 2)
Model 2: number_flowers ~ poly(week, 2) + PC1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1          93    80.503
2          92    78.699 1    1.8042 0.1792
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 8d: Output del test ANOVA che confronta il modello ridotto (Model 1) e il modello completo (Model 2)

Anche dai test ANOVA emerge che i fattori del meteo nei giorni pregressi all'osservazione non incidono in alcun modo sul numero di fiori presenti in una piantina allocata in qualsiasi campo su cui è stato installato un impianto agrivoltaico; invece le piantine situate nel campo con condizione di luce piena (FL) sono sensibili al meteo che potrebbe alterarne la quantità.

## COMMENTO

Dalle analisi la prima componente principale si è ritenuta adeguata e sufficiente per spiegare i dati in quanto coglie gran parte della varianza totale in ciascuna condizione di luce.

Il test *ANOVA* e l'osservazione del coefficiente che interessa la componente determinano se i fattori meteo dei giorni antecedenti alle rilevazioni sulle piantine influenzano l'evoluzione delle piantine e delle sue caratteristiche oggetto di studio.

Nelle condizioni luce FL e SH30, ossia nei campi maggiormente esposti alla luce solare e agli agenti metereologici in generale, l'altezza e le foglie, nello specifico la loro lunghezza e la quantità, sono sensibili al meteo dei giorni precedenti alla rilevazione.

Il numero di foglie e la quantità di fiori appaiono influenzati significativamente dai fattori del meteo rispettivamente nelle piantine allocate nel campo soggetto a luce piena (FL) e in HSH60; condizione di luce in cui il meteo influisce anche sul diametro secondario.

Dalle analisi si conclude che un campo caratterizzato da un impianto agrivoltaico SH60 imprime sul terreno una notevole porzione d'ombra, dunque copre in modo tale da proteggere le piantine da eventuali agenti metereologici estremi e da un'eccessiva esposizione ai raggi solari che potrebbero danneggiarle.





## CONCLUSIONI

In seguito a un'analisi grafica svolte sulle variabili che rappresentano le caratteristiche oggetto di studio delle piantine di patate, in generale un impianto agrivoltaico permette delle performance leggermente migliori rispetto a quelle riscontrate nella condizione di luce piena (FL).

Modelli polinomiali sono stati affiancati a modelli GAM per verificare se questi sono in grado di cogliere un andamento più vicino ai dati, ma il confronto non ha riscontrato differenze significative tra i due modelli selezionati per ciascuna variabile studiata; in ciascun modello si sono notati andamenti con valori più elevati nelle piantine situate in campi contraddistinti da un impianto.

Si è inoltre deciso di verificare se mediamente le particolarità delle piantine cambiano in base alla condizione di luce a cui sono sottoposte, ossia se la suddivisione in campi determina differenze in termini di crescita delle piantine. A eccezione dei due diametri, lo sviluppo delle peculiarità mutano mediamente all'interno dei diversi campi osservati, i due diametri invece presentano una variazione simile in tutte le piantine soggette ad analisi. Infine si è considerato anche il meteo e dalle analisi appare che i fattori meteo influenzano l'altezza, le foglie e i fiori delle piantine, mentre lo sviluppo dei due diametri non sono sensibili al meteo dei giorni precedenti. Un campo sottoposto a una maggiore esposizione e quindi parzialmente coperto o privo della struttura, permette alle piantine di essere influenzate dal meteo, invece un campo in cui vi è installato un impianto più imponente protegge le piantine rendendo queste non influenzabili dai agenti meteorologici.

L'agrivoltaico comporta vantaggi: garantisce la protezione delle culture dal Sole e dagli eventi meteorologici estremi, riduce l'evaporazione aumentando l'umidità del suolo e diminuisce la temperatura del terreno nelle giornate calde limitando così il consumo di acqua per irrigare; un sistema agrivoltaico è pertanto favorevole per la coltivazione in generale. Per la cultura trattata, un impianto apporta chiaramente dei lievi miglioramenti e benefici alla coltivazione rispetto a una condizione di luce piena. Tra gli impianti agrivoltaici, SH60 genera sul campo un ombreggiamento medio costante pari al 60% dell'intera area e dalle analisi risulta che una quantità così elevata d'ombra non giova alla coltivazione di patate.



## BIBLIOGRAFIA

- Amaducci, S., Yin, X., & Colauzzi, M. (2018). Agrivoltaic systems to optimise land use for electric energy production. *Applied Energy*, 220, 545-561.  
<https://doi.org/10.1016/j.apenergy.2018.03.081>
- Coiante, D. (2008). *Fotovoltaico. Il processo evolutivo e le nuove frontiere*. ENEA
- Linee Guida in materia di Impianti Agrivoltaici. (2022b, June). *Ministero Dell'Ambiente E Della Sicurezza Energetica*. <https://www.mase.gov.it/>
- Turano, V. (2023, January 16). *Agenda 2030 per lo sviluppo sostenibile*. Agenzia per La Coesione Territoriale. <https://www.agenziacoesione-gov.it/comunicazione/agenda-2030-per-lo-sviluppo-sostenibile/>
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)* (2nd ed.). Chapman and Hall/CRC.
- Johnson, R. A., & Wichern, D. W. (2013). *Applied Multivariate Statistical Analysis: Pearson New International Edition*. Pearson Education.
- Grigoletto, M., Ventura, L., & Pauli, F. (2017). *Modello Lineare: Teoria e Applicazioni con R*. Giappichelli.
- Bruno, A. A. W. G. S. (2012). *Data Analysis and Data Mining: An Introduction*. Oxford University Press, USA.
- Faraway, J. J. (2005). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. CRC Press.