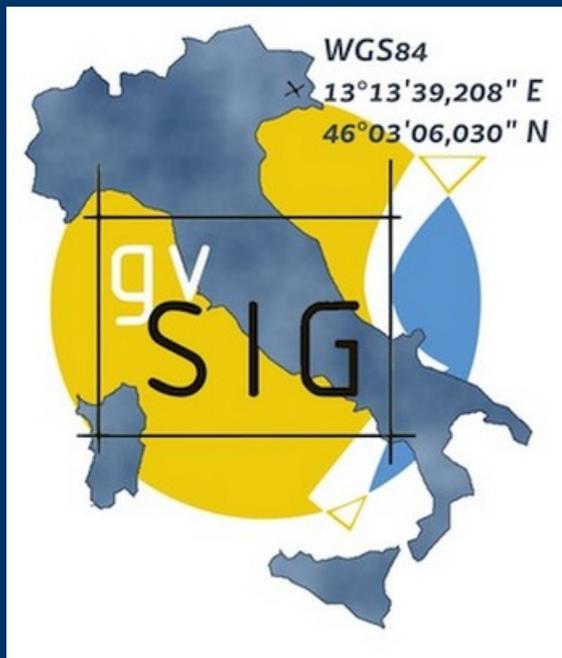




Università degli studi di Udine

Fabio Crosilla

Dimitri Macorig



Comune di Tavagnacco

Ivano Sebastianutti

QUARTE GIORNATE ITALIANE DI gvSIG

Classificazione dei dati laser scanning mediante analisi statistica dei momenti di ordine superiore della distribuzione dei punti: prime applicazioni al sistema informativo territoriale del Comune di Tavagnacco

Udine, 21 aprile 2011

INTRODUZIONE

Lo studio verte su un'originale applicazione metodologica relativa ad una delle più innovative tecniche di rilevamento del territorio: il laser scanning

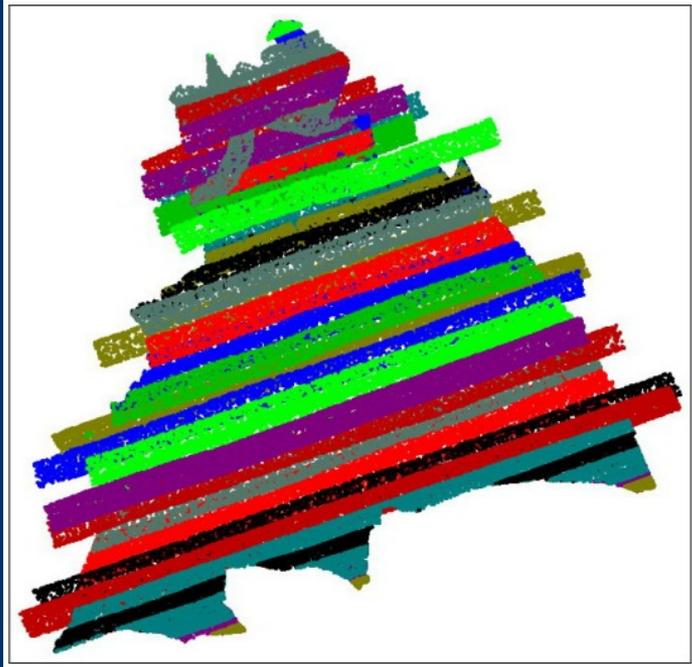
Lo studio associa il lavoro di classificazione dei dati laser, svolto presso il Comune di Tavagnacco (UD), alla più attuale ricerca bibliografica ed entra nel merito della classificazione di tipo statistico dei punti laser

Ciò ha permesso di identificare un metodo di classificazione basato sul calcolo degli indici di asimmetria e curtosi, affiancato ad un'analisi della distribuzione di punti laser sia per i valori di intensità radiometrica che per altezza, consentendo l'utilizzo del metodo in modo iterativo ed interattivo

Si è quindi reso necessario lo sviluppo di un programma per il trattamento dei dati affiancato ad una sperimentazione su casi reali presenti nel territorio comunale di Tavagnacco

RILIEVO LASER E DATI DI OUTPUT

Il **Comune di Tavagnacco** ha svolto nel 2007 un rilievo aereofotogrammetrico affiancato ad un rilievo con laser scanner dell'intero territorio comunale ottenuto con il sistema **Leica ALS50**



Altezza di volo: **1000 m**

Densità: **12 punti/mq**



Si sono ottenute 40 strisciate costituite da altrettanti file di output in formato testuale (.XYZ) che vengono convertiti in formato .las per un più facile utilizzo

| | | | | |
|----|------------|-------------|---------|----|
| 1 | 362253.450 | 5110801.190 | 209.600 | 75 |
| 2 | 362253.490 | 5110801.110 | 209.580 | 71 |
| 3 | 362253.530 | 5110801.030 | 209.580 | 67 |
| 4 | 362253.570 | 5110800.950 | 209.570 | 76 |
| 5 | 362253.610 | 5110800.870 | 209.570 | 76 |
| 6 | 362253.650 | 5110800.780 | 209.540 | 84 |
| 7 | 362253.690 | 5110800.700 | 209.540 | 69 |
| 8 | 362253.730 | 5110800.620 | 209.540 | 67 |
| 9 | 362253.770 | 5110800.540 | 209.530 | 73 |
| 10 | 362253.810 | 5110800.460 | 209.530 | 65 |

Lo scanner rileva, per ciascun punto, oltre alle tre coordinate nel sistema di riferimento WGS84, il valore dell'intensità radiometrica (con un valore compreso tra 0 e 255)

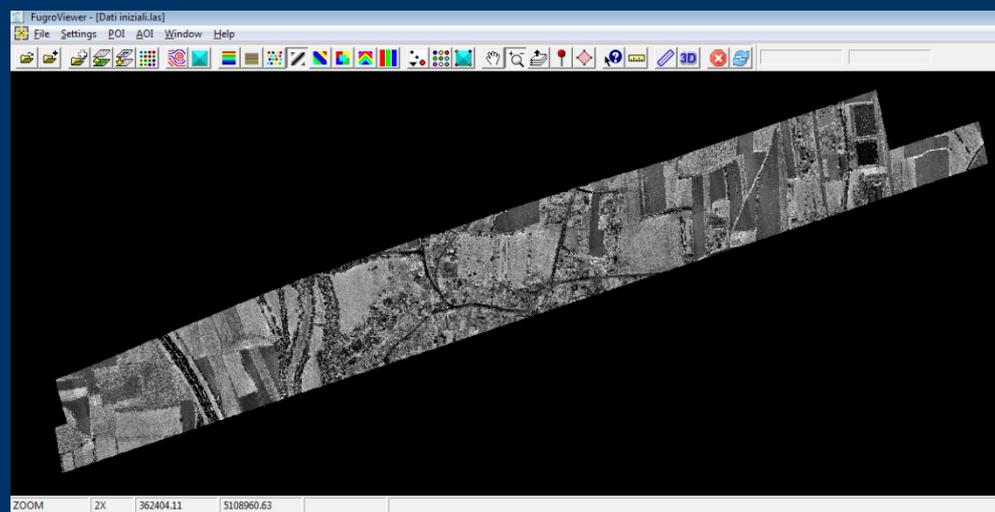
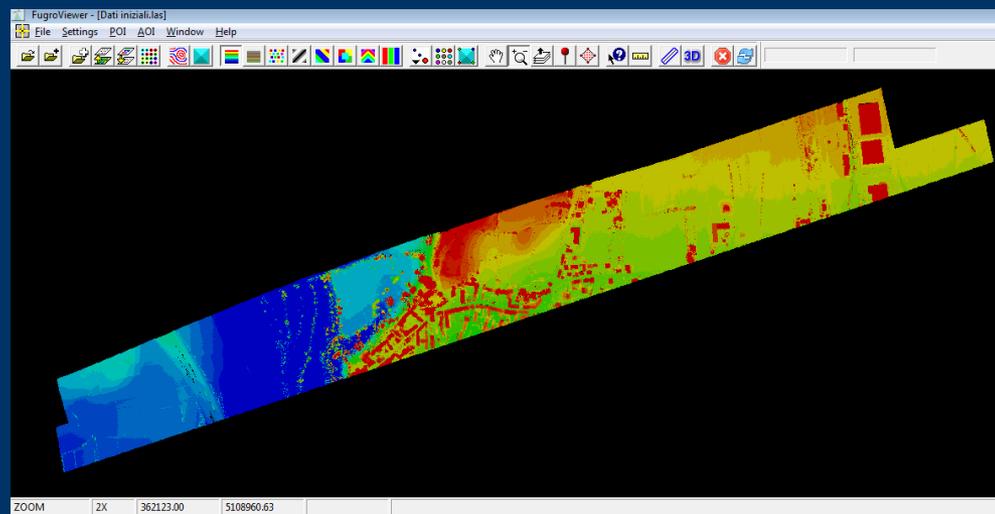
CLASSIFICAZIONE PRELIMINARE

Si è scelta una zona del Comune su cui iniziare una prima classificazione di tipo manuale per avere dei dati di riferimento correttamente classificati; per fare ciò si sono importati i file della zona in formato .las (con **lastools**)

Si possono visualizzare tutti i dati forniti dal rilievo laser

- Visualizzazione per altezza

- Visualizzazione per intensità

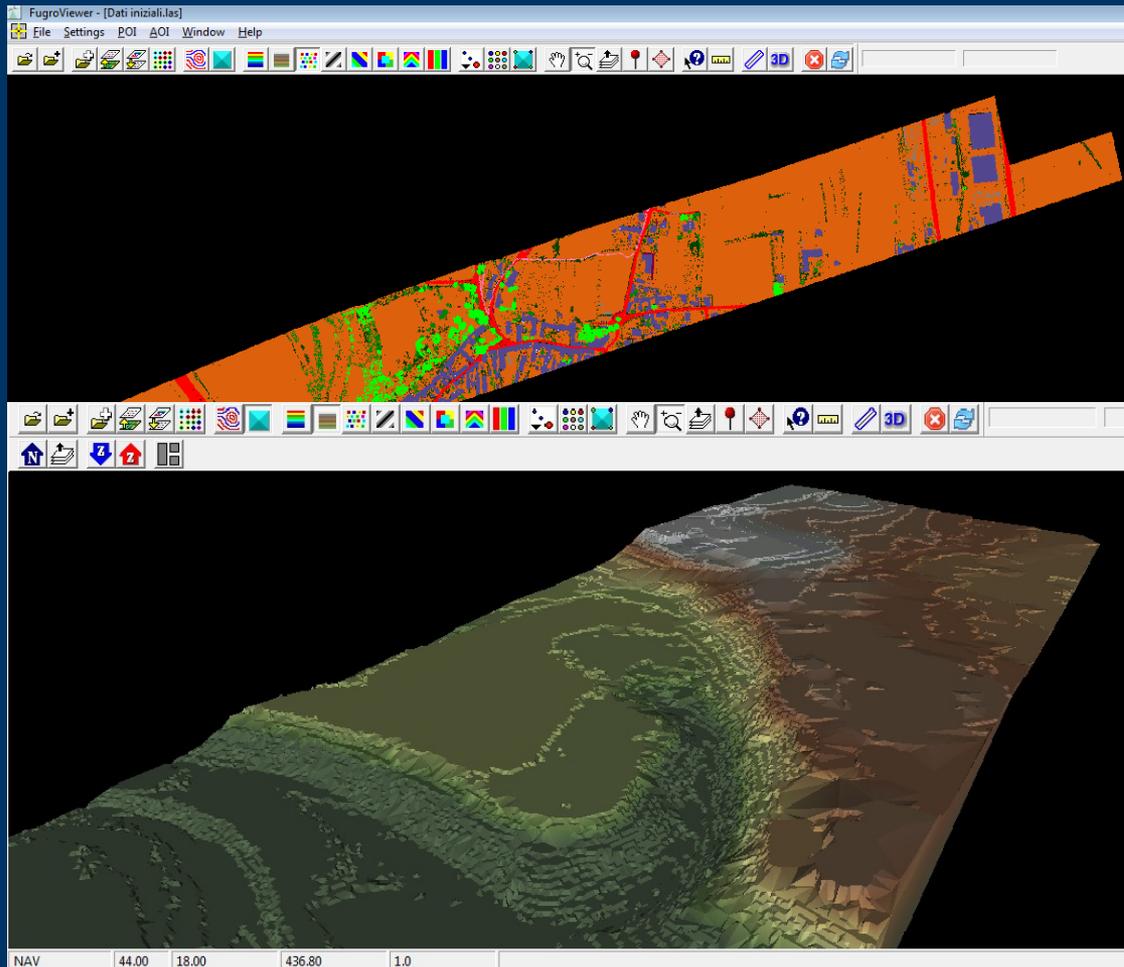


CLASSIFICAZIONE DEI PUNTI

I punti sono stati classificati in 11 famiglie identificate da un numero di riferimento che viene aggiunto al file come quinta colonna

Il tutto lo si può visualizzare in funzione dei diversi colori per ogni classe

| | | |
|----|--------------------------|--|
| 0 | Punti iniziali | |
| 1 | Punti non classificati | |
| 2 | Terreno | |
| 3 | Bassa vegetazione | |
| 4 | Media vegetazione | |
| 5 | Alta vegetazione | |
| 6 | Edifici | |
| 14 | Strade | |
| 15 | Arredo urbano | |
| 16 | Pista ciclabile | |
| 17 | Punti sopra la copertura | |



| X | Y | Z | I | C |
|-------------|---------------|---------|----|---|
| 360.806,547 | 5.109.570,97 | 187,976 | 43 | 2 |
| 360.806,567 | 5.109.570,90 | 187,944 | 44 | 2 |
| 360.806,588 | 5.109.570,83 | 187,922 | 43 | 2 |
| 360.806,613 | 5.109.570,72 | 188,004 | 46 | 2 |
| 360.806,636 | 5.109.570,61 | 188,064 | 57 | 2 |
| 360.806,655 | 5.109.570,56 | 187,933 | 43 | 2 |
| 360.806,677 | 5.109.570,501 | 187,950 | 55 | 2 |
| 360.806,695 | 5.109.570,463 | 187,847 | 43 | 2 |
| 360.806,718 | 5.109.570,365 | 187,889 | 42 | 2 |
| 360.806,743 | 5.109.570,255 | 187,953 | 44 | 2 |
| 360.806,765 | 5.109.570,171 | 187,966 | 66 | 2 |
| 360.806,787 | 5.109.570,097 | 187,944 | 42 | 2 |
| 360.806,810 | 5.109.570,015 | 187,938 | 43 | 2 |
| 360.806,830 | 5.109.569,955 | 187,887 | 37 | 2 |
| 360.806,854 | 5.109.569,867 | 187,893 | 36 | 2 |
| 360.806,878 | 5.109.569,769 | 187,928 | 38 | 2 |
| 360.806,900 | 5.109.569,682 | 187,947 | 40 | 2 |
| 360.806,924 | 5.109.569,595 | 187,947 | 45 | 2 |
| 360.806,947 | 5.109.569,499 | 187,976 | 65 | 2 |
| 360.806,969 | 5.109.569,426 | 187,947 | 39 | 2 |
| 360.806,993 | 5.109.569,342 | 187,941 | 42 | 2 |
| 360.807,017 | 5.109.569,240 | 187,988 | 44 | 2 |
| 360.807,039 | 5.109.569,168 | 187,953 | 39 | 2 |
| 360.807,063 | 5.109.569,077 | 187,964 | 37 | 2 |
| 360.807,086 | 5.109.569,000 | 187,940 | 39 | 2 |
| 360.807,109 | 5.109.568,914 | 187,940 | 48 | 2 |
| 360.807,134 | 5.109.568,809 | 187,986 | 39 | 2 |

Sono stati classificati più di un milione di punti, l'80% dei quali appartengono al terreno, l'8% c.a. alla vegetazione, il 7% c.a. all'edificato ed il 3% c.a. alle strade

DEFINIZIONI UTILI PER COMPRENDERE IL METODO DI CLASSIFICAZIONE STATISTICO (1)

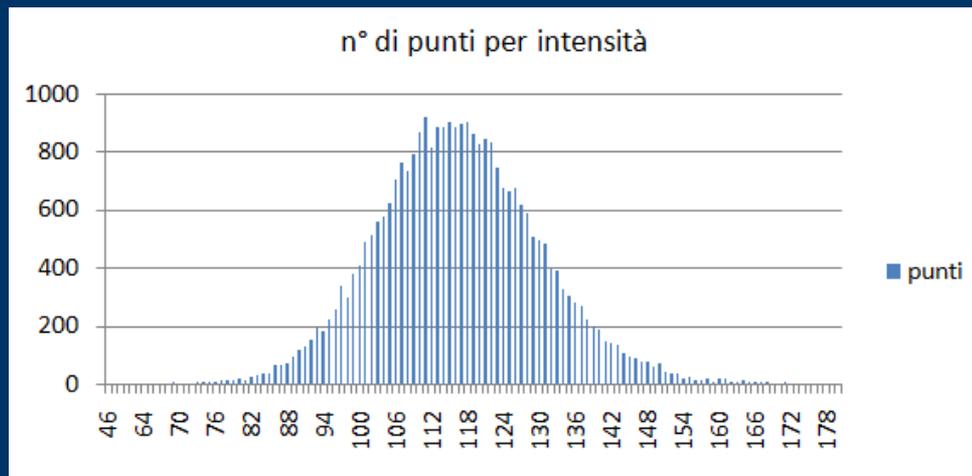
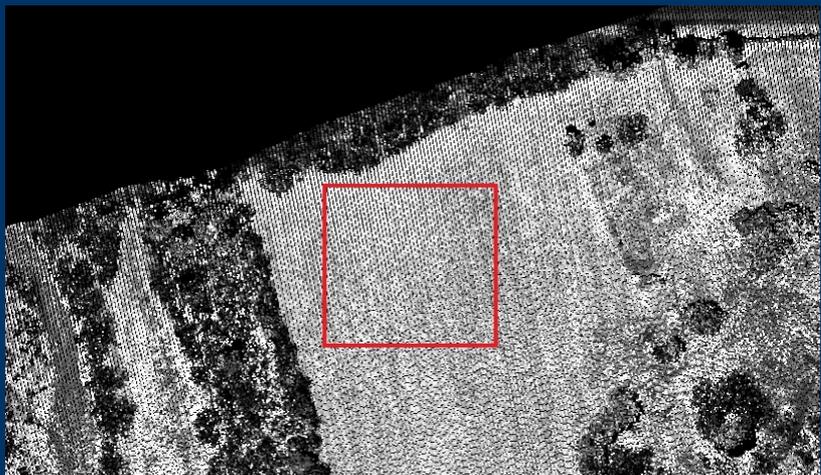
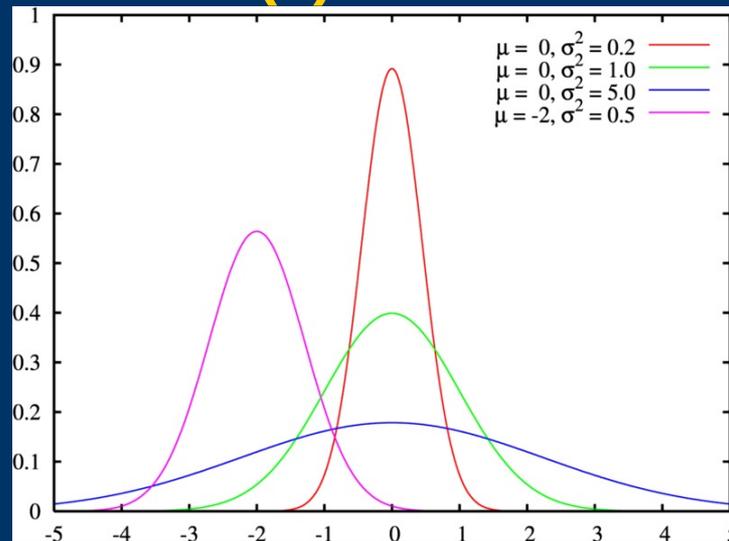
Distribuzione normale: distribuzione di probabilità caratterizzata dalla seguente funzione di densità:

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\mu = \frac{\sum x_i}{N}$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

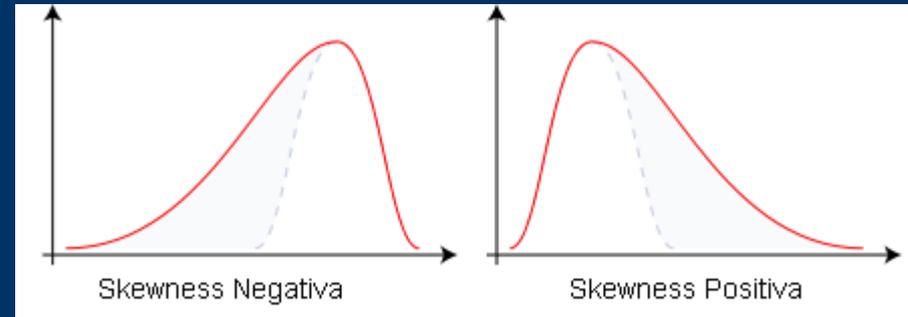
In particolar modo questa caratteristica si è riscontrata analizzando i dati relativi all'intensità del rilievo laser su una zona di terreno omogenea



DEFINIZIONI UTILI PER COMPRENDERE IL METODO DI CLASSIFICAZIONE STATISTICO (2)

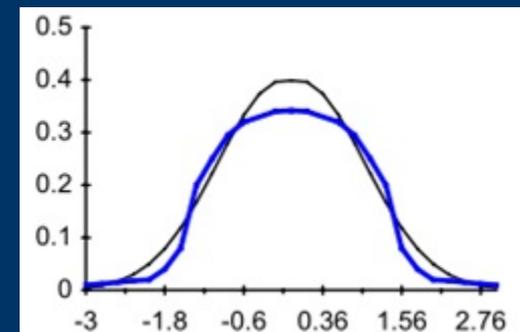
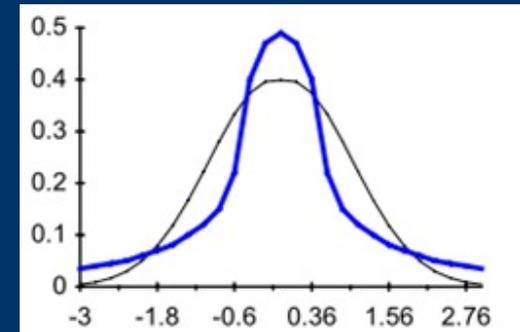
Esistono due indici che consentono di evidenziare quanto una distribuzione si allontani dalla normalità:

• L'**indice di asimmetria** (skewness in inglese) rappresenta quanto il picco della distribuzione si discosti dal picco della normale; il suo valore può essere negativo, positivo o pari a zero (perfetta simmetria)



$$Y_1 = \frac{m_3}{\sigma^3} = \left(\frac{1}{N \times \sigma^3} \times \sum_{i=1}^N (x_i - \mu)^3 \right)^{1/3}$$

• L'**indice di curtosi** che rappresenta il grado di appiattimento rispetto ad una distribuzione normale; se il valore è > 3 la distribuzione è **leptocurtica**; se il valore è < 3 la distribuzione è **platicurtica**



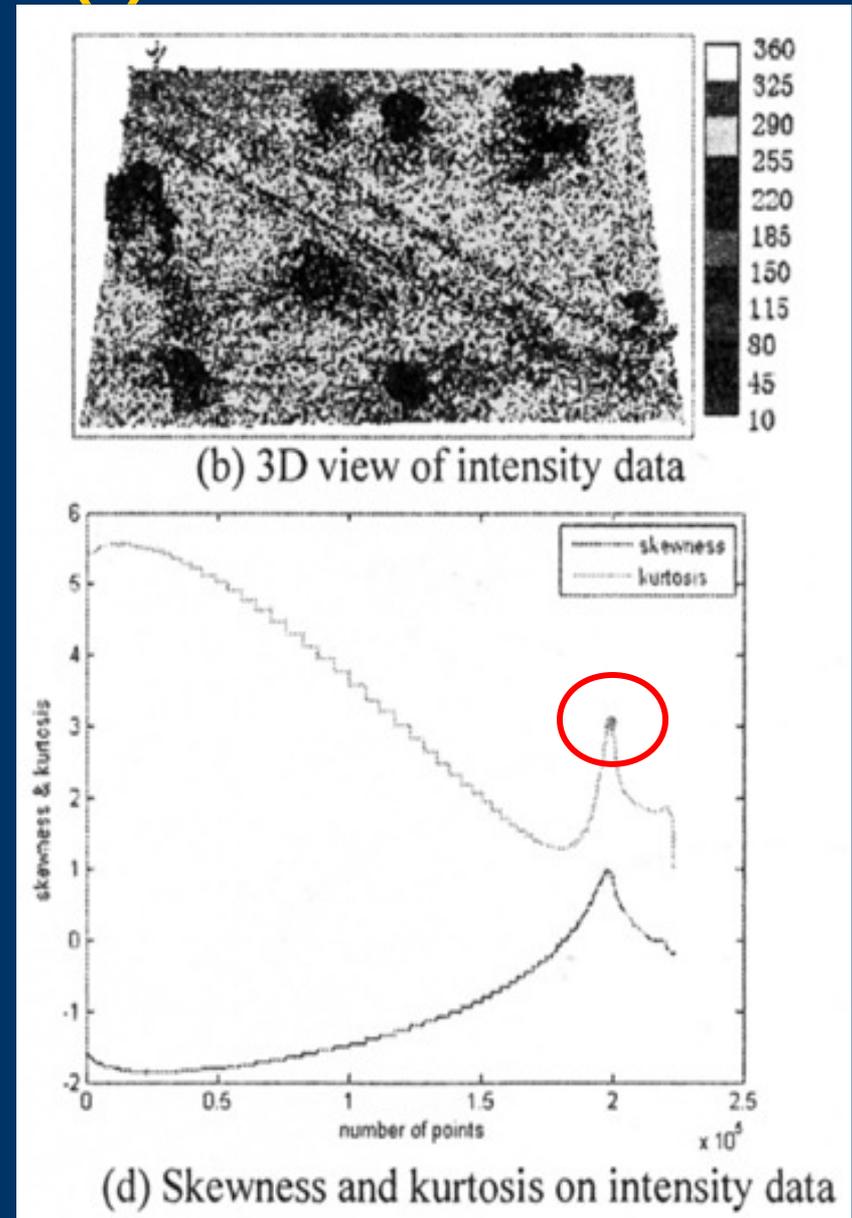
$$Y_2 = \frac{m_4}{\sigma^4} = \left(\frac{1}{N \times \sigma^4} \times \sum_{i=1}^N (x_i - \mu)^4 \right)^{1/4}$$

ANALISI STATISTICA CON GLI INDICI DI ASIMMETRIA E CURTOSI (2)

In Liu et al. (2009) si fa un confronto utilizzando sia il metodo relativo all'intensità che quello riferito all'altezza classificando i punti terreno in corrispondenza di punti di massimo relativo di curtosi

Ragionando per intensità si eliminano dal calcolo i punti con valore più elevato (punti bianchi), ripetendo il calcolo delle funzioni in modo iterativo si ottiene l'andamento in figura

Il picco di massimo relativo si avrà verso la fine delle iterazioni, esaltando la famiglia dei punti scuri (punti albero); oltre il massimo relativo di curtosi si classifica ciò che sta sopra il terreno

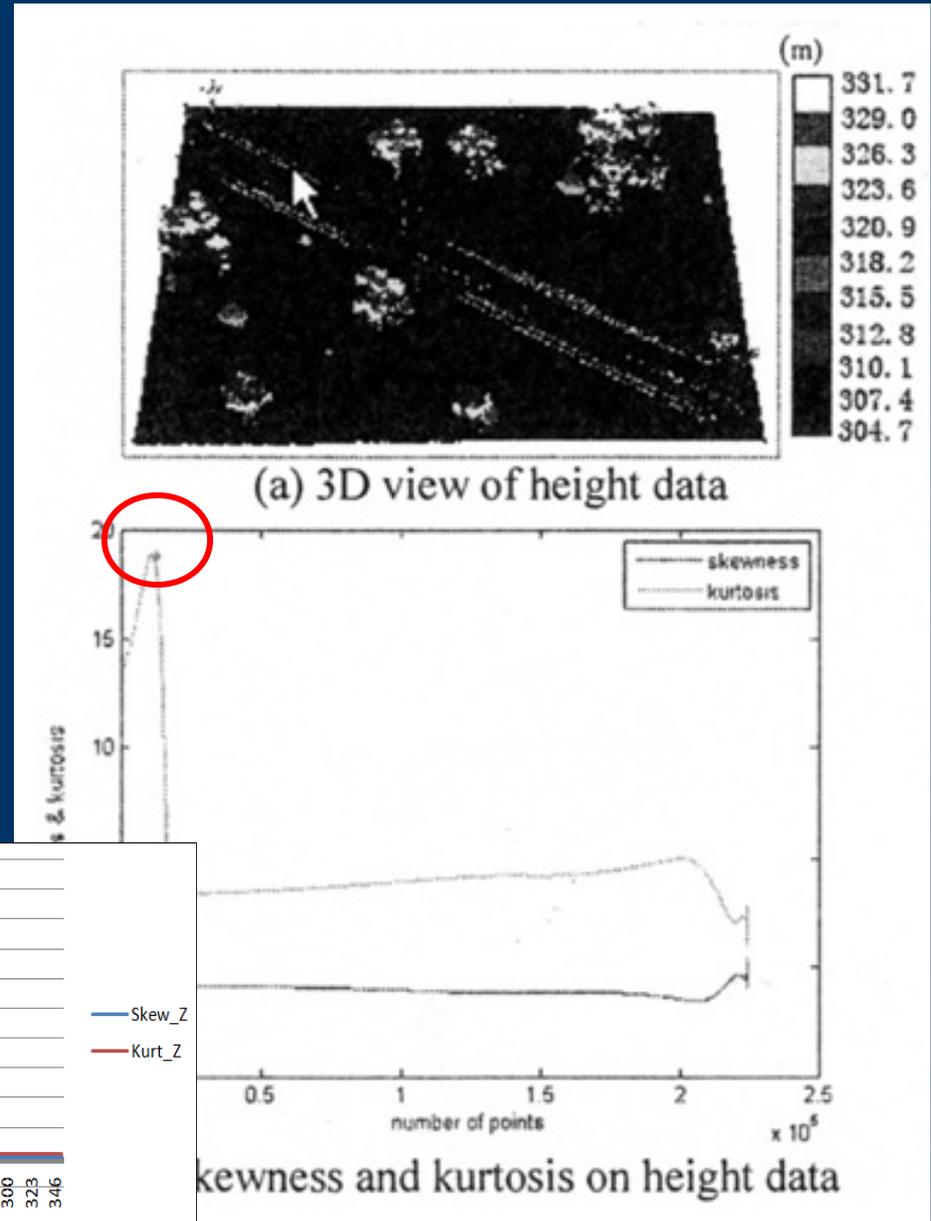


ANALISI STATISTICA CON GLI INDICI DI ASIMMETRIA E CURTOSI (3)

Ragionando per **altezza**, nel calcolo iterativo si eliminano prima i punti a quota maggiore (alberi) che sono in numero inferiore rispetto ai punti terreno

Si ottiene quindi un picco di curtosi nelle prime iterazioni; i punti dopo tale picco sono tutti punti terreno

Si è però reso necessario effettuare ulteriori sperimentazioni e si è visto che la corretta classificazione si ha all'appiattimento della funzione di curtosi



CREAZIONE DEL PROGRAMMA DI CLASSIFICAZIONE (1)

Il programma consiste in un file che viene trattato con un software per la gestione di database (es. **MsAccess**, **postgresql**, **MySQL**, ecc.) che riceve di input il file relativo alla zona da analizzare che è costituito dalle coordinate del punto (X,Y,Z) nelle prime tre colonne, i valori di intensità (I) nella quarta colonna e nella quinta la classificazione manuale del punto (C)

| X | Y | Z | I | C |
|-------------|---------------|---------|----|---|
| 360.806,547 | 5.109.570,971 | 187,974 | 43 | 2 |
| 360.806,567 | 5.109.570,906 | 187,944 | 44 | 2 |
| 360.806,588 | 5.109.570,835 | 187,922 | 43 | 2 |
| 360.806,613 | 5.109.570,721 | 188,004 | 45 | 2 |
| 360.806,636 | 5.109.570,619 | 188,064 | 57 | 2 |
| 360.806,655 | 5.109.570,566 | 187,994 | 54 | 2 |
| 360.806,677 | 5.109.570,501 | 187,954 | 55 | 2 |
| 360.806,695 | 5.109.570,463 | 187,844 | 43 | 2 |
| 360.806,718 | 5.109.570,365 | 187,885 | 42 | 2 |
| 360.806,743 | 5.109.570,255 | 187,953 | 44 | 2 |
| 360.806,765 | 5.109.570,171 | 187,964 | 65 | 2 |
| 360.806,787 | 5.109.570,097 | 187,944 | 42 | 2 |
| 360.806,810 | 5.109.570,015 | 187,933 | 43 | 2 |
| 360.806,830 | 5.109.569,955 | 187,887 | 37 | 2 |
| 360.806,854 | 5.109.569,867 | 187,893 | 35 | 2 |
| 360.806,878 | 5.109.569,769 | 187,923 | 38 | 2 |
| 360.806,900 | 5.109.569,682 | 187,947 | 40 | 2 |
| 360.806,924 | 5.109.569,595 | 187,947 | 45 | 2 |
| 360.806,947 | 5.109.569,499 | 187,974 | 65 | 2 |
| 360.806,969 | 5.109.569,426 | 187,947 | 39 | 2 |
| 360.806,993 | 5.109.569,342 | 187,947 | 42 | 2 |
| 360.807,017 | 5.109.569,240 | 187,983 | 44 | 2 |
| 360.807,039 | 5.109.569,168 | 187,953 | 39 | 2 |
| 360.807,063 | 5.109.569,077 | 187,964 | 37 | 2 |
| 360.807,086 | 5.109.569,000 | 187,944 | 39 | 2 |
| 360.807,109 | 5.109.568,914 | 187,944 | 43 | 2 |
| 360.807,134 | 5.109.568,809 | 187,984 | 39 | 2 |

Si utilizzano tutta una serie di query create per ottenere nella tabella Area_laser i dati relativi al calcolo degli indici per tutti i punti presenti nel file; ciò può essere fatto sia per i valori di intensità che per quelli geometrici

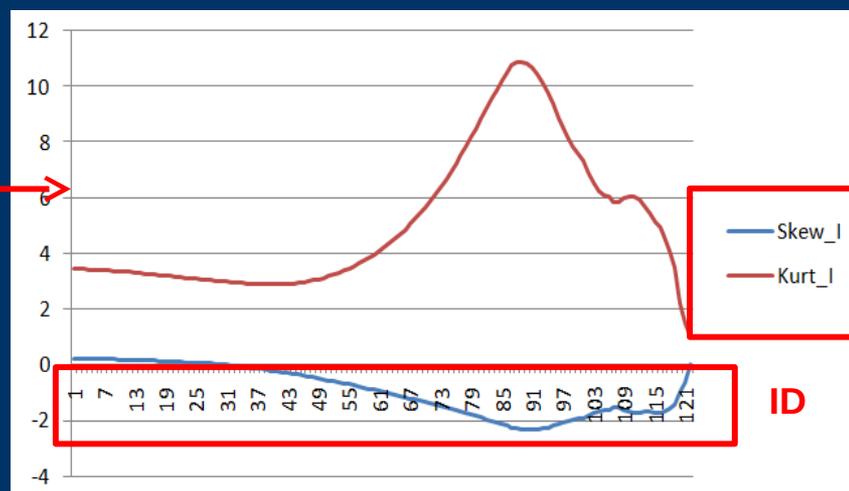
| Nome campo | Tipo dati |
|------------|-----------|
| X | Numerico |
| Y | Numerico |
| Z | Numerico |
| I | Numerico |
| C | Numerico |
| FlagZ | Si/No |
| FlagI | Si/No |
| NC_I | Numerico |
| NC_Z | Numerico |
| Skew_I | Numerico |
| Kurt_I | Numerico |
| Skew_Z | Numerico |
| Kurt_Z | Numerico |

CREAZIONE DEL PROGRAMMA DI CLASSIFICAZIONE (2)

Contemporaneamente questi valori vengono riportati in una tabella (valori_I o valori_Z) in cui ad ogni iterazione è associato un numero identificativo (ID) a cui corrispondono i valori di asimmetria e curtosi; questo viene poi tradotto in un grafico

| Nome campo | Tipo dati |
|------------|-----------|
| ID | Contatore |
| N°_I | Numerico |
| Varianza_I | Numerico |
| Skew_I | Numerico |
| Kurt_I | Numerico |
| Media_I | Numerico |
| Max_I | Numerico |
| Max_Z | Numerico |
| Min_Z | Numerico |

| Nome campo | Tipo dati |
|------------|-----------|
| X | Numerico |
| Y | Numerico |
| Z | Numerico |
| I | Numerico |
| C | Numerico |
| FlagZ | Si/No |
| FlagI | Si/No |
| NC_I | Numerico |
| NC_Z | Numerico |
| Skew_I | Numerico |
| Kurt_I | Numerico |
| Skew_Z | Numerico |
| Kurt_Z | Numerico |

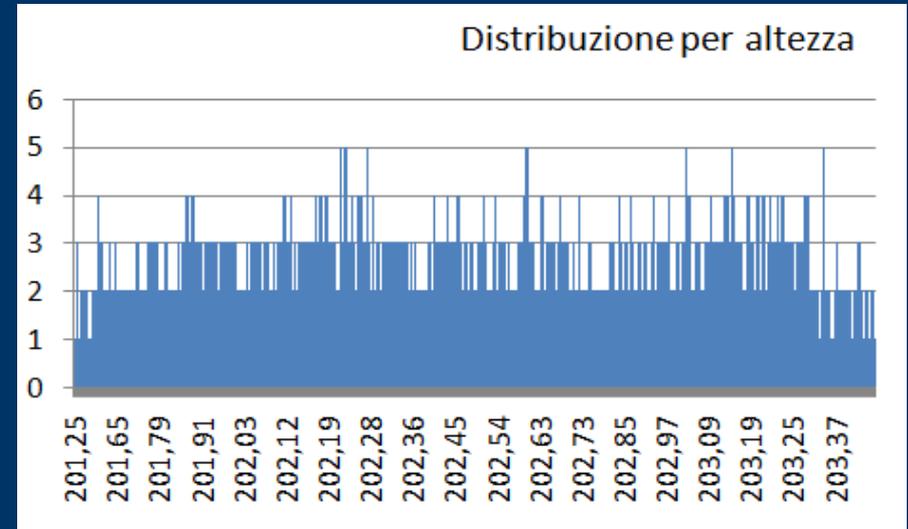
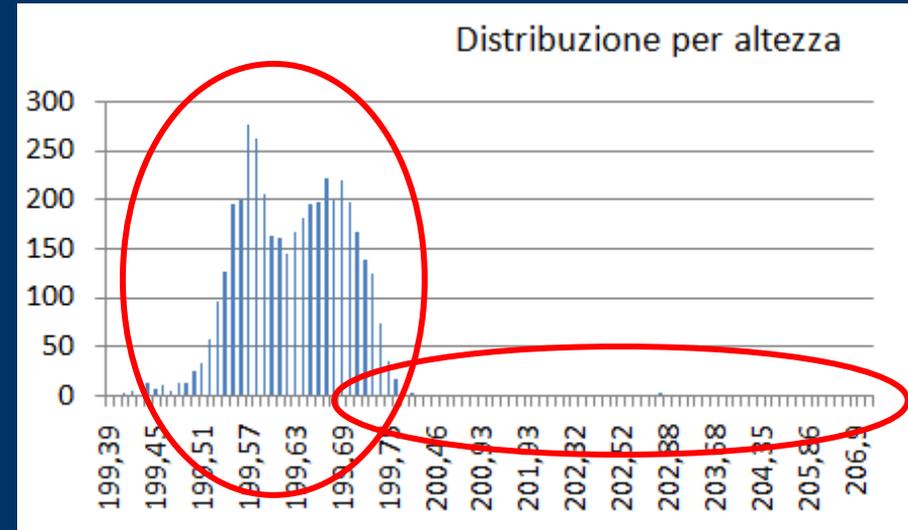
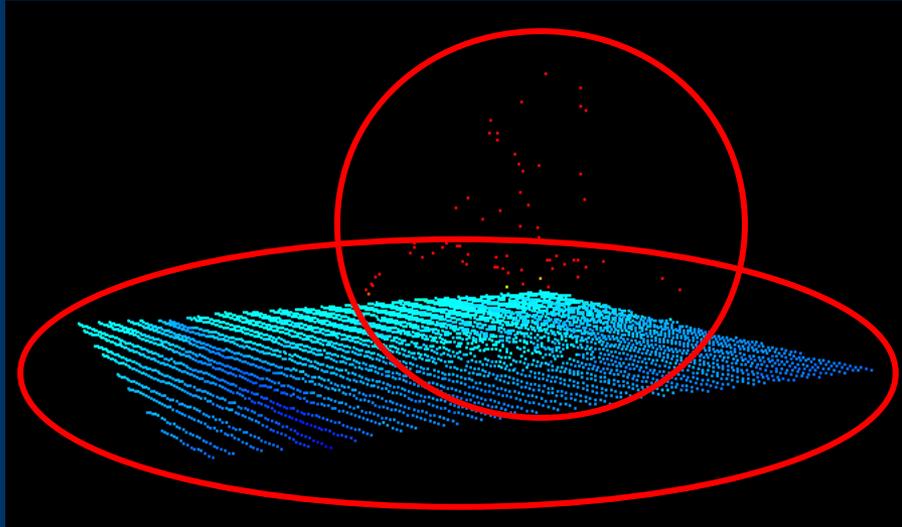


Una volta deciso a quale ciclo si intende classificare i punti laser, lo si immette in un'apposita query (Imposta_Kurt) e si associa al n° di ciclo un codice di classificazione. L'output che si ottiene è un file contenente X,Y,Z,I,C opportunamente riconvertito in formato .las

STIMA NON-PARAMETRICA DELLA DENSITÀ (1)

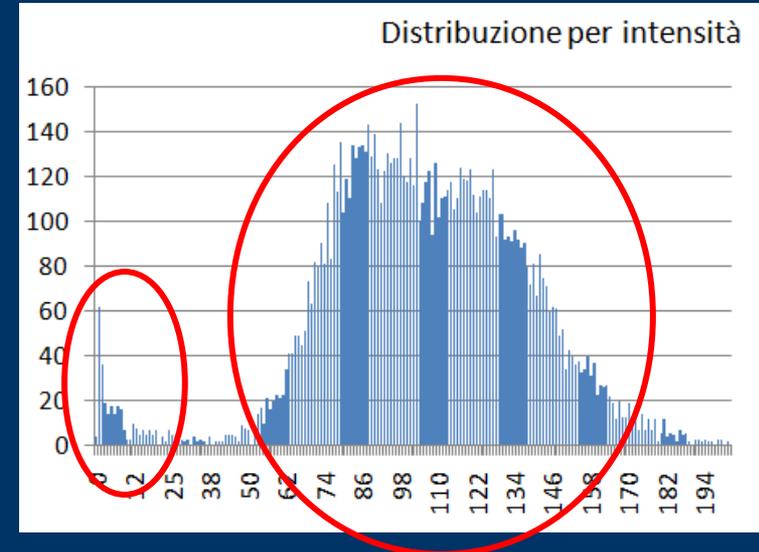
Si riportano alcuni esempi riguardanti la distribuzione dei punti per altezza ed intensità riscontrati nelle varie sperimentazioni

Valutazione dell'altezza



STIMA NON-PARAMETRICA DELLA DENSITÀ (2)

Valutazione dell'intensità



Si rende necessario l'utilizzo di un metodo rigoroso per la determinazione delle diverse famiglie di punti presenti nella distribuzione

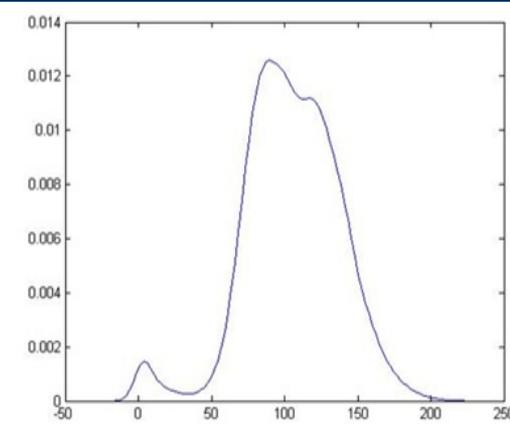
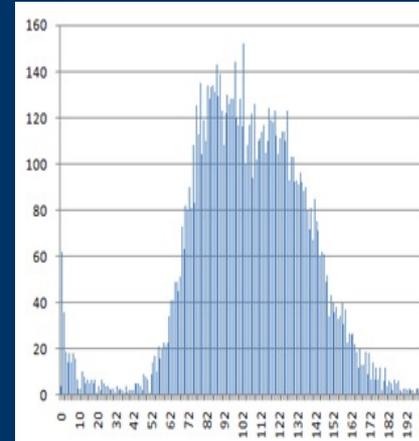
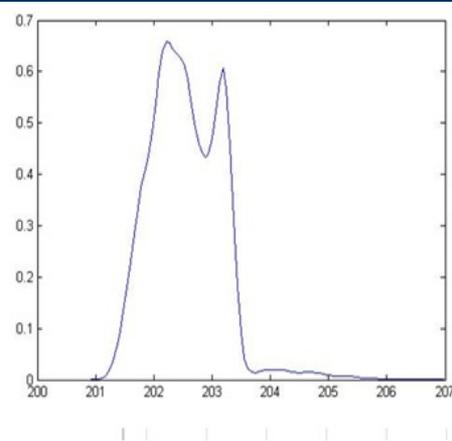
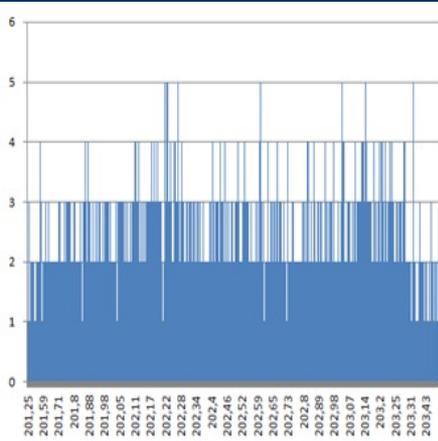
STIMA NON-PARAMETRICA DELLA DENSITÀ (3)

Il criterio non-parametrico è indipendente dalla distribuzione originale dalla quale proviene il campione

Ad ogni valore campionato si assegna una distribuzione di probabilità detta “kernel” (K); in particolare noto un campione (x_1, x_2, \dots, x_n) preso da una distribuzione avente densità f incognita, si vuole stimare la forma di questa funzione f dalle relazioni:

$$f_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) \quad \text{con} \quad K_h = \frac{1}{h} K\left(\frac{x}{h}\right)$$

Con l'utilizzo di un software statistico (es. **R**) è possibile effettuare questo calcolo:

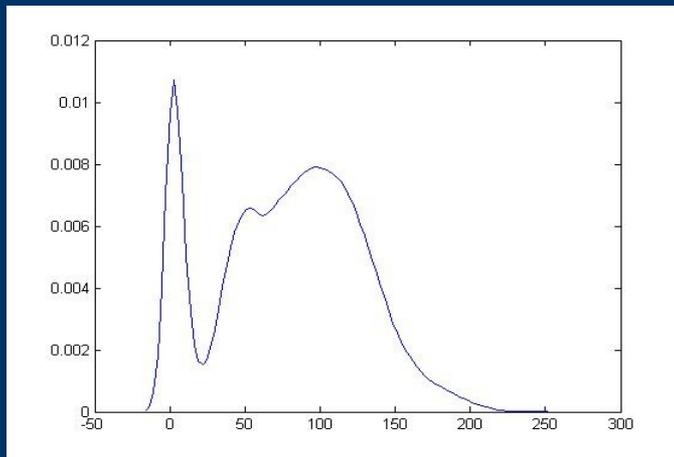


DIP TEST (HARTIGAN 1985)

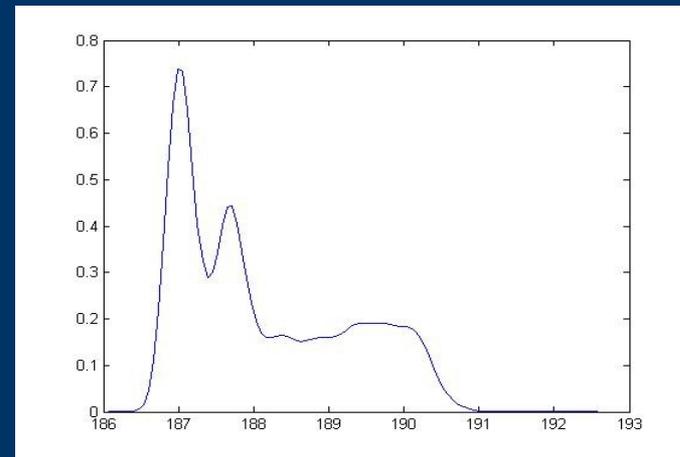
Nell'ottica dell'automazione del programma si è cercato di sostituire il controllo manuale dei diagrammi delle distribuzioni; ciò ha portato all'utilizzo del dip test, tale da evidenziare quanto una distribuzione è unimodale o multi-modale

Utilizzando il programma **R** si richiama dalla libreria il dip test e si ottiene un valore che è tanto più grande quanto è marcata la sua multi-modalità

```
> I<-as.matrix(read.table("I.txt"))
> library(diptest)
> dip(I)
[1] 0.0542717
> |
```

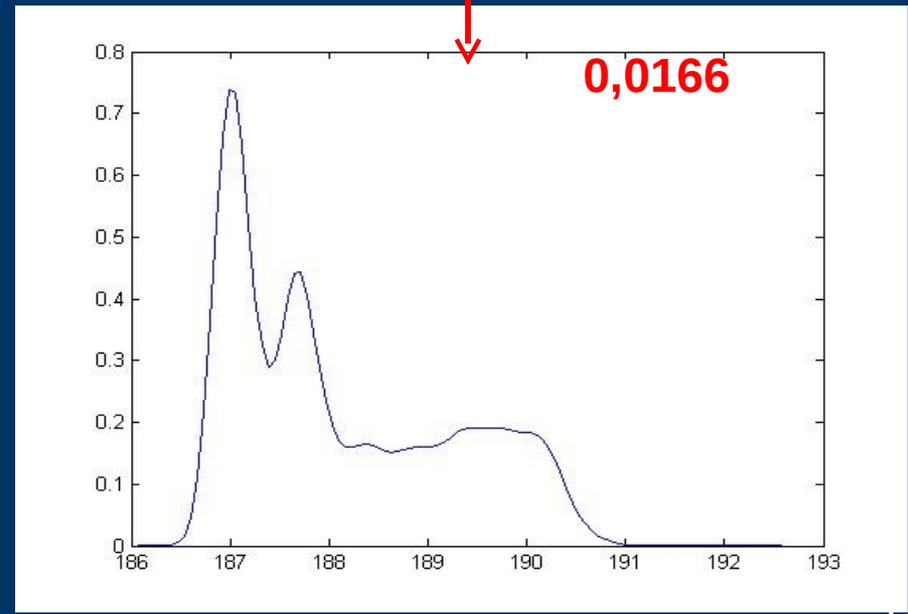
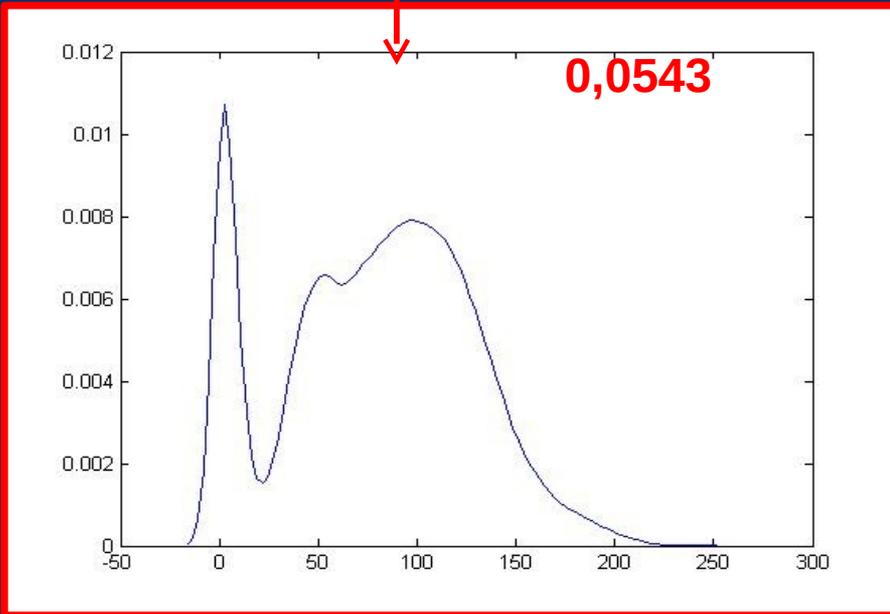
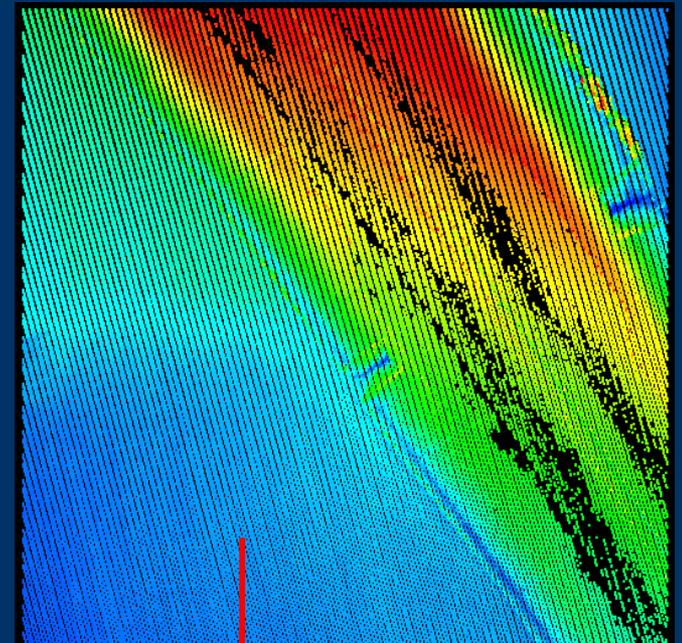
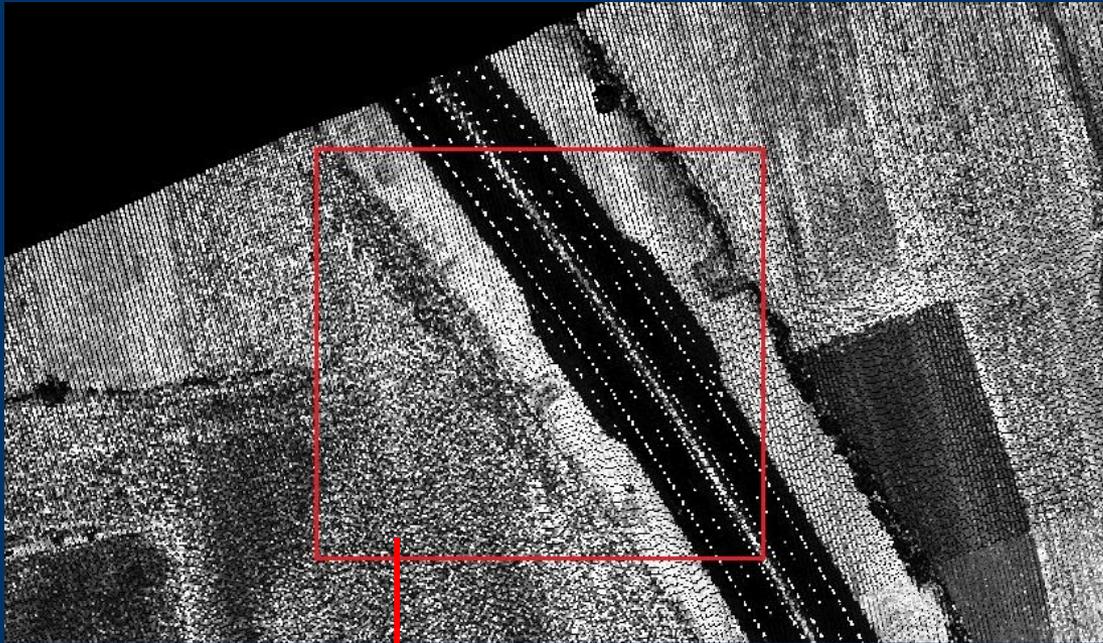


0,0543

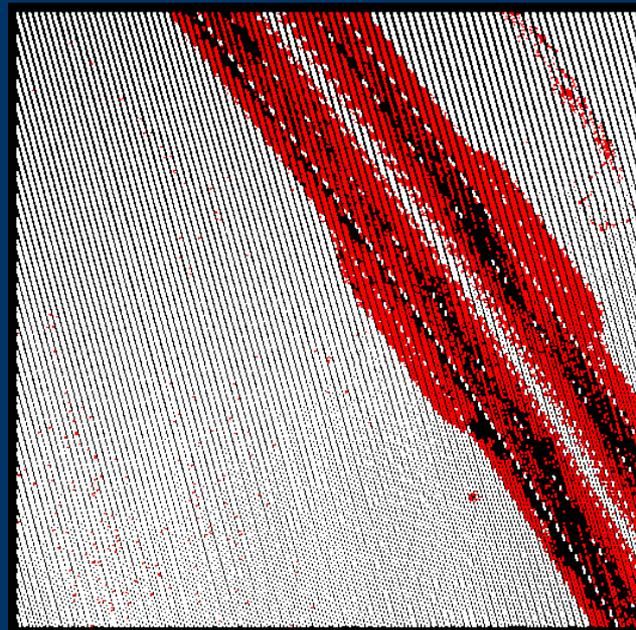
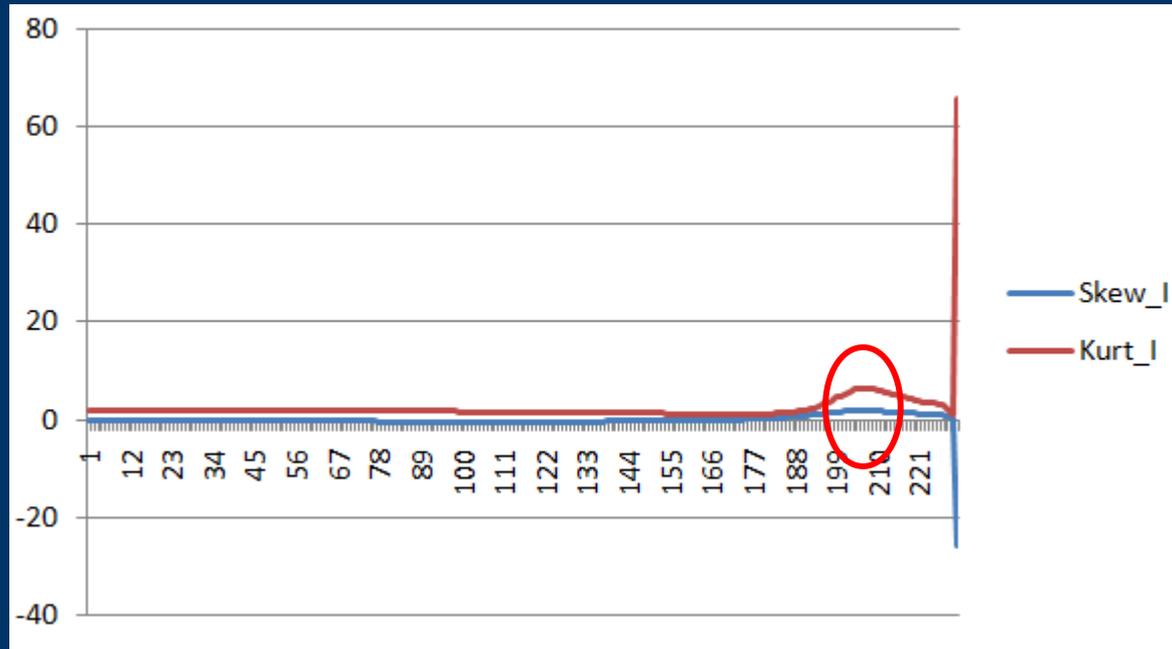


0,0166

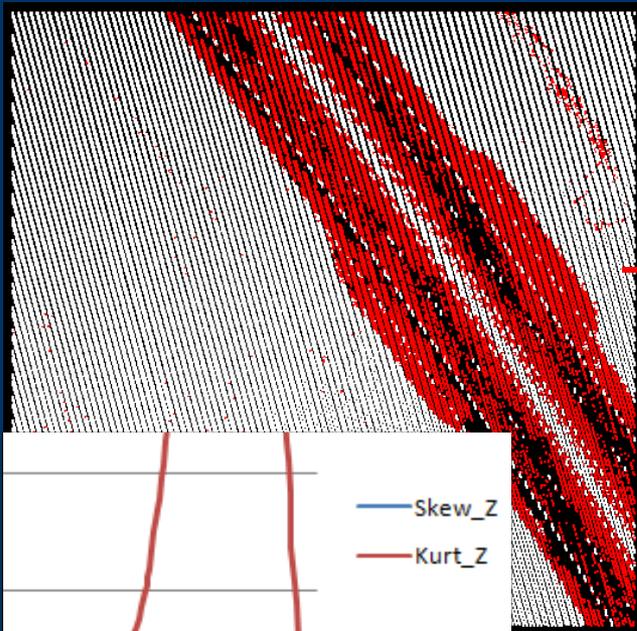
PRIMO ESPERIMENTO (1)



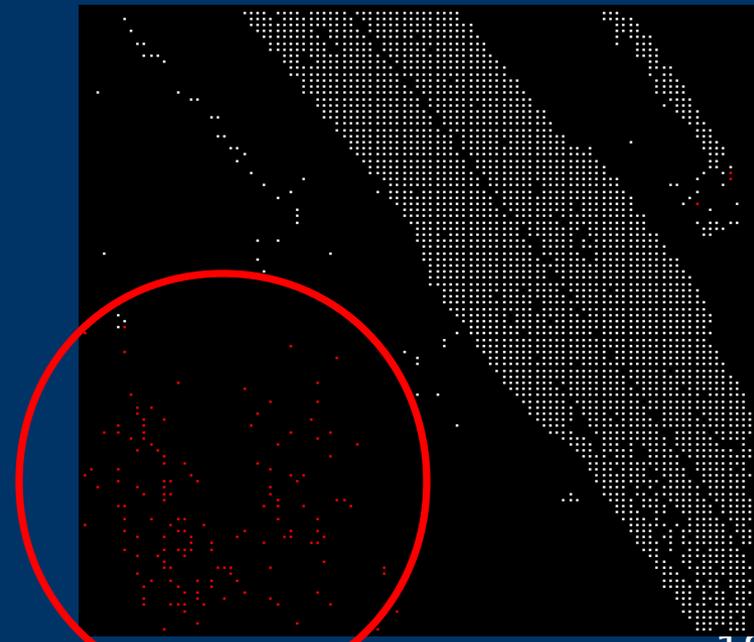
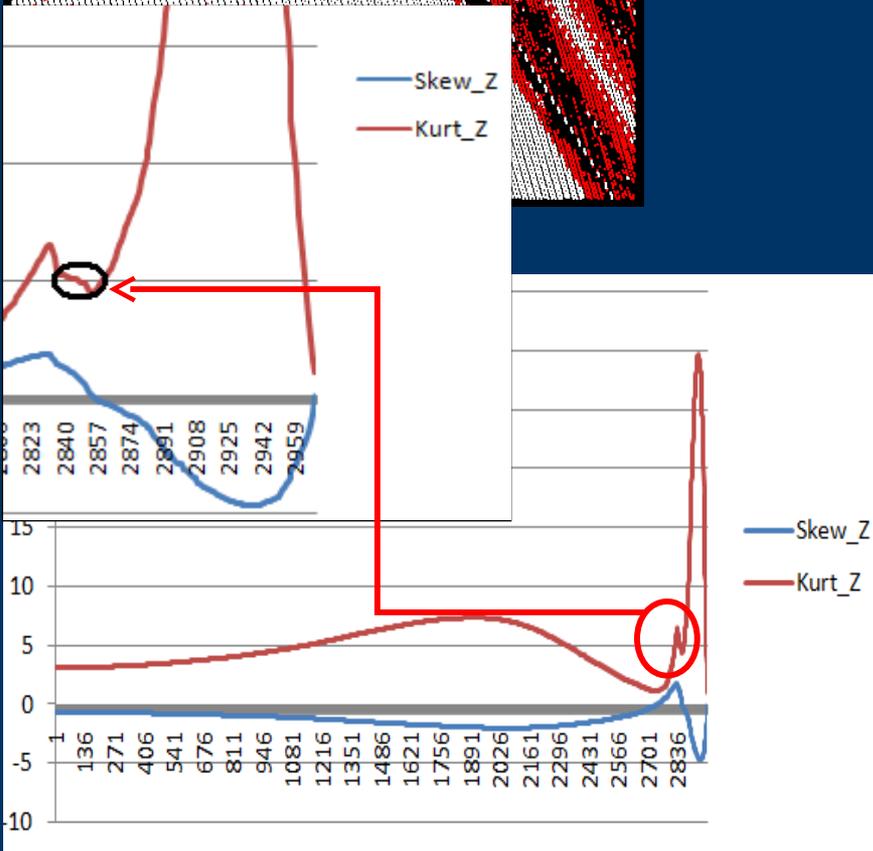
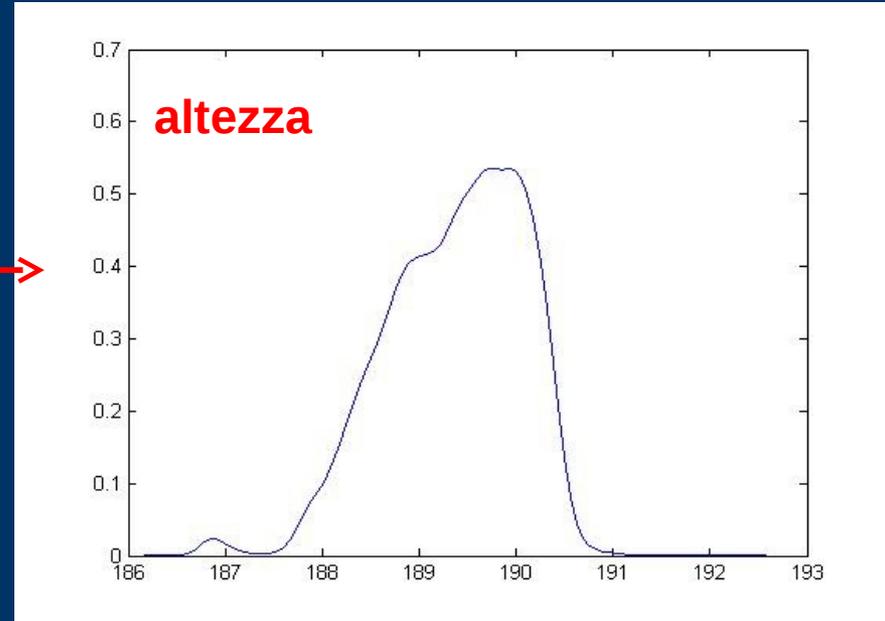
PRIMO ESPERIMENTO (2)



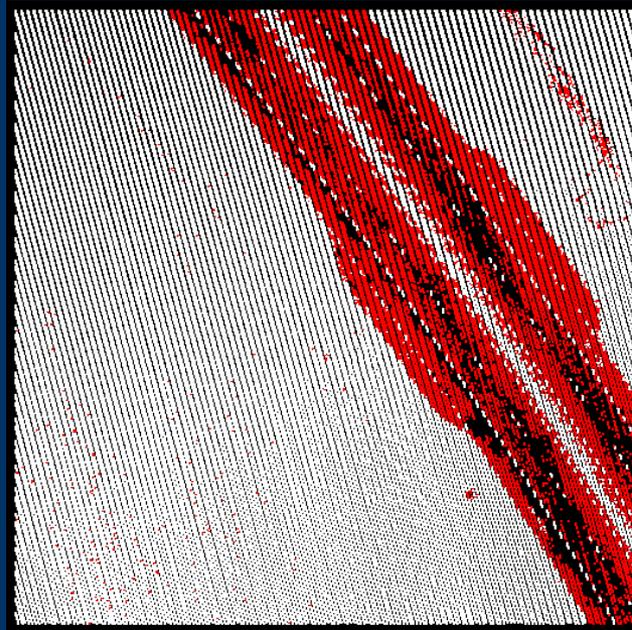
PRIMO ESPERIMENTO (3)



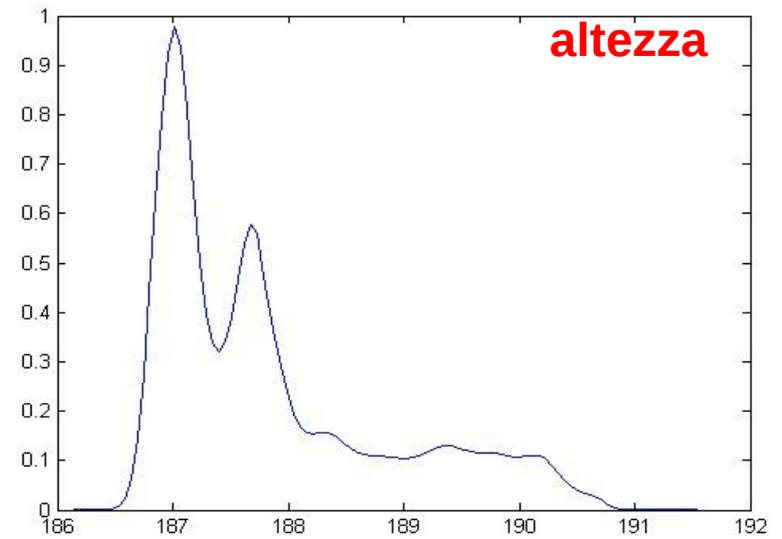
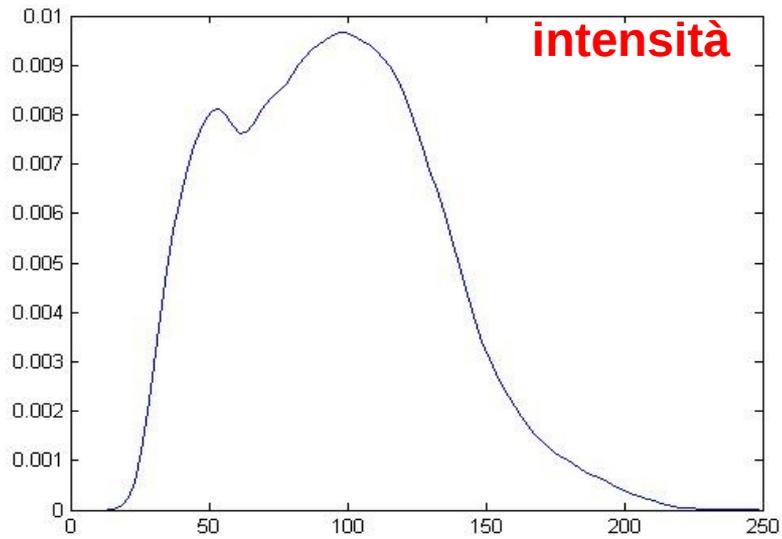
Valutazione punti rossi



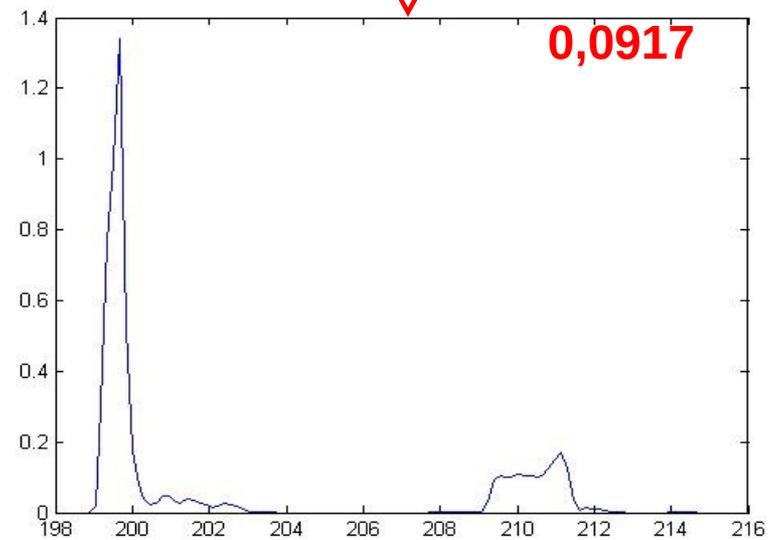
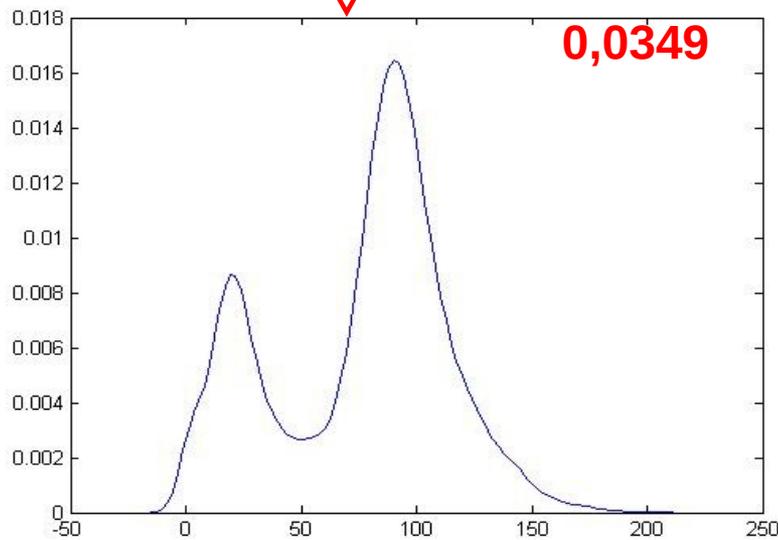
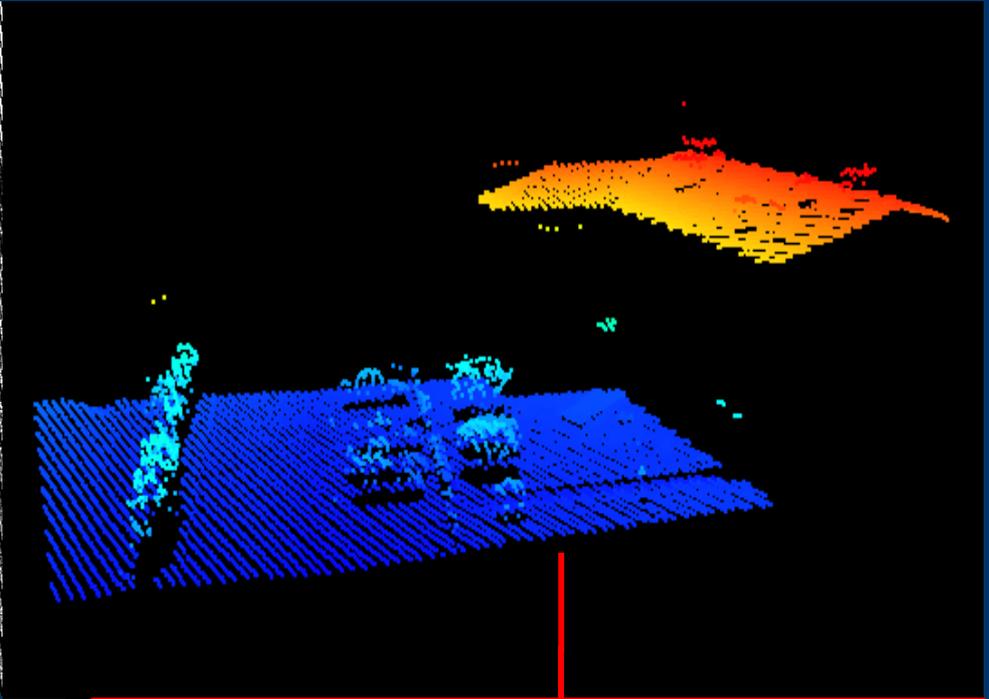
PRIMO ESPERIMENTO (4)



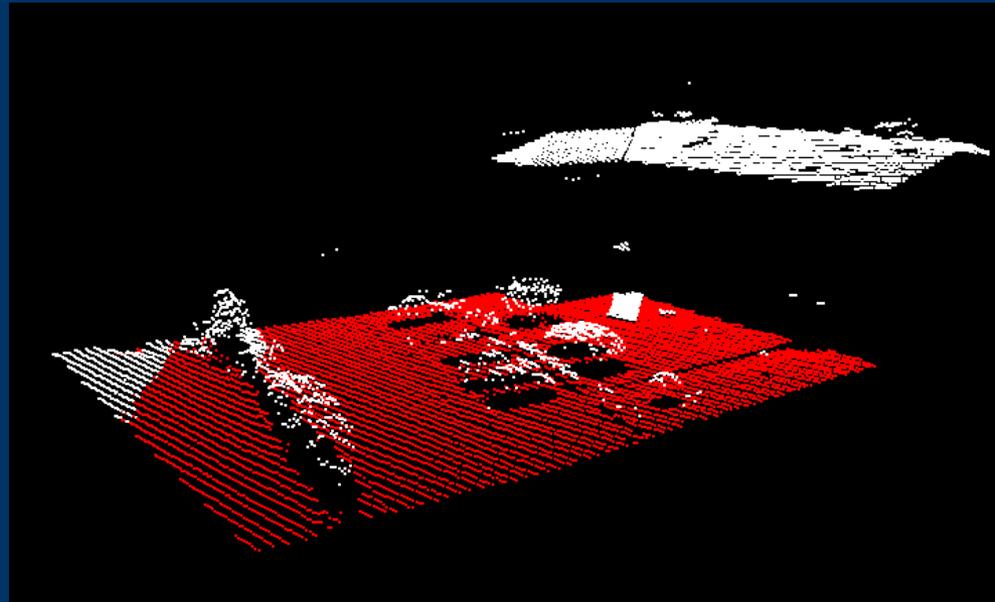
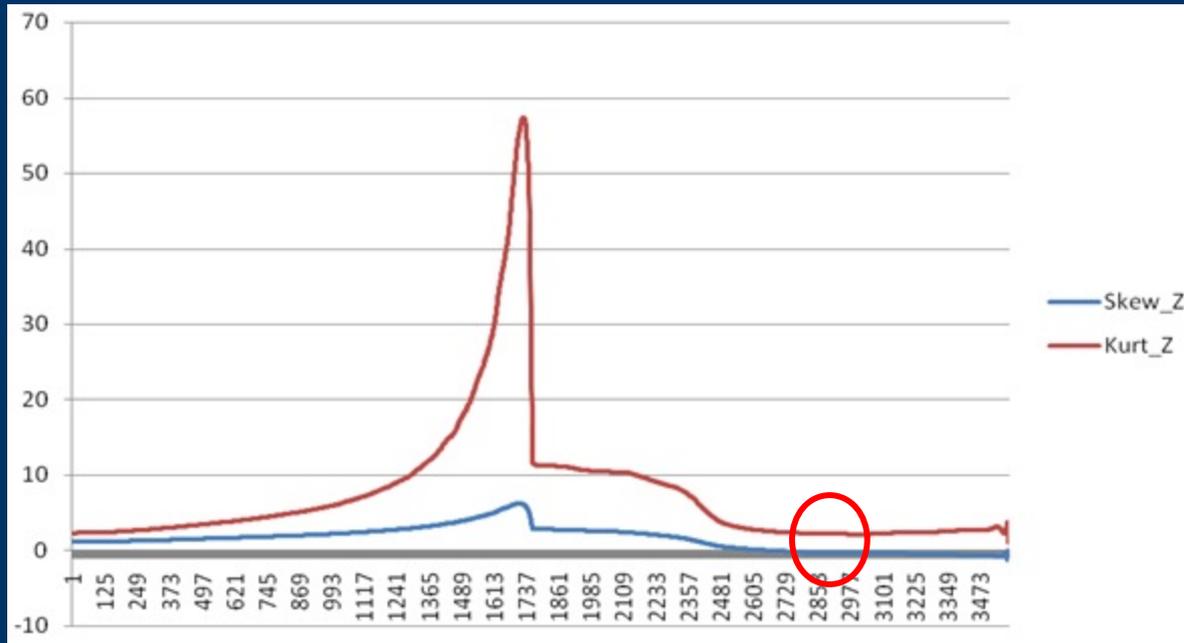
Valutazione punti
bianchi



SECONDO ESPERIMENTO (1)

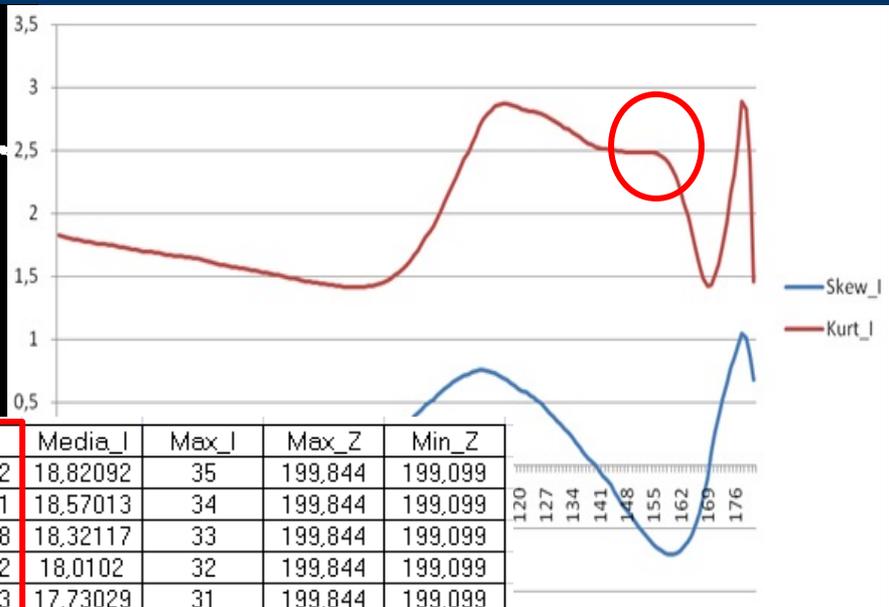
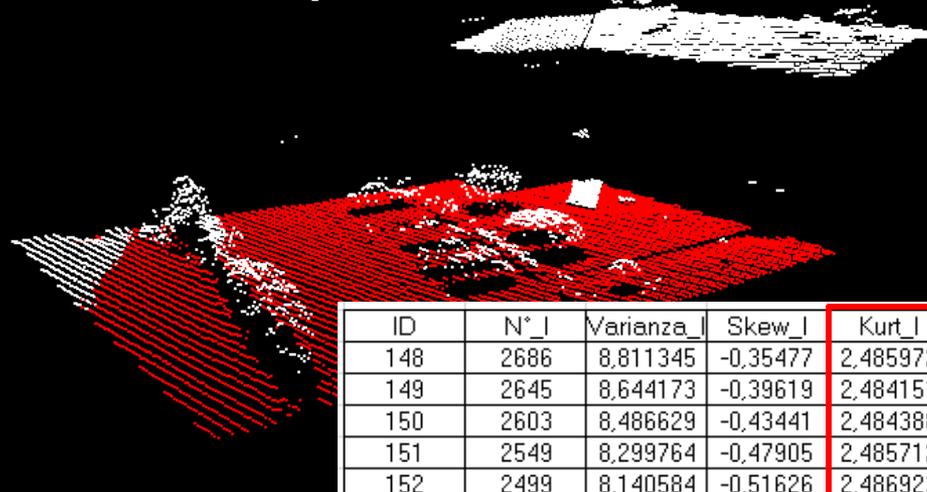


SECONDO ESPERIMENTO (2)

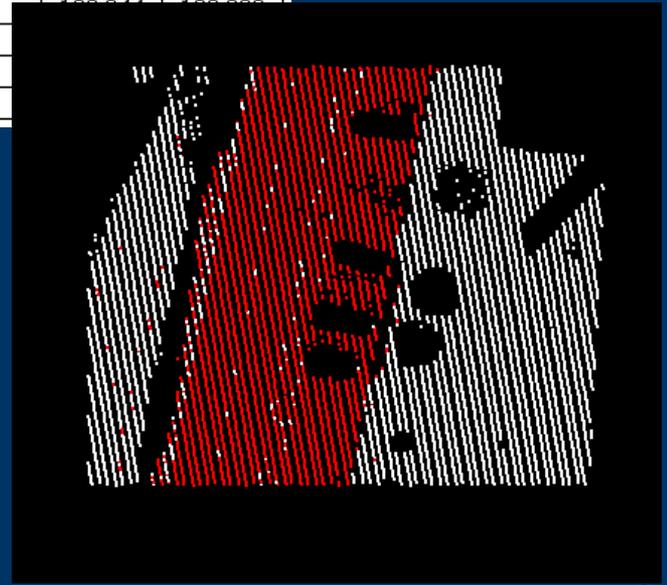
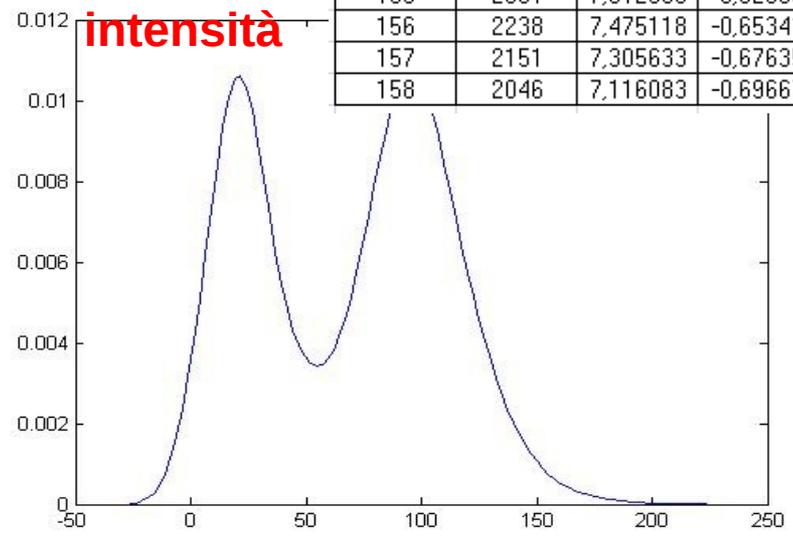


SECONDO ESPERIMENTO (3)

Valutazione punti rossi

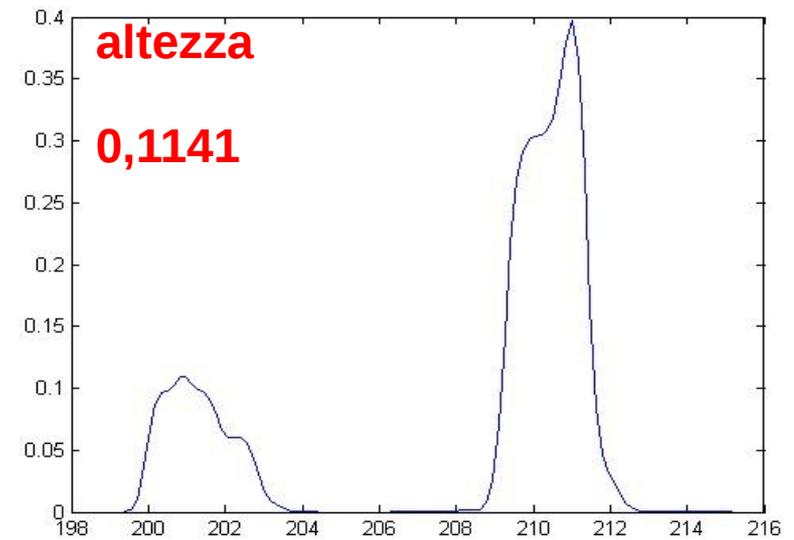
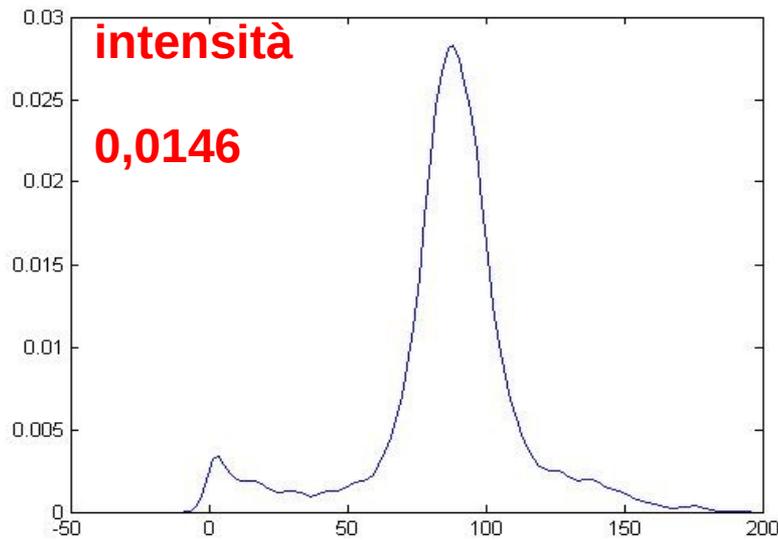
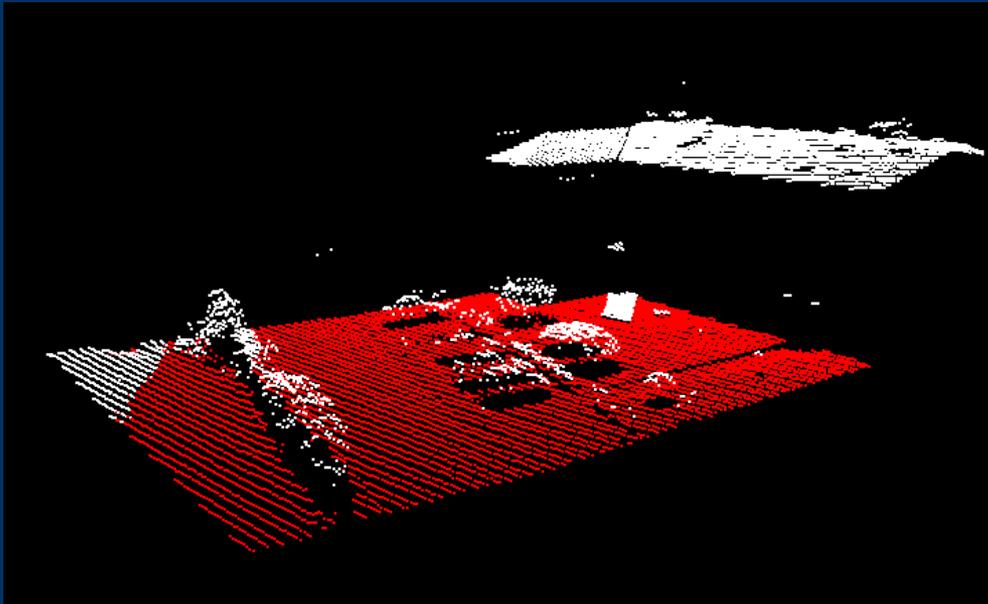


| ID | N°_I | Varianza_I | Skew_I | Kurt_I | Media_I | Max_I | Max_Z | Min_Z |
|-----|------|------------|----------|----------|----------|-------|---------|---------|
| 148 | 2686 | 8,811345 | -0,35477 | 2,485972 | 18,82092 | 35 | 199,844 | 199,099 |
| 149 | 2645 | 8,644173 | -0,39619 | 2,484151 | 18,57013 | 34 | 199,844 | 199,099 |
| 150 | 2603 | 8,486629 | -0,43441 | 2,484388 | 18,32117 | 33 | 199,844 | 199,099 |
| 151 | 2549 | 8,299764 | -0,47905 | 2,485712 | 18,0102 | 32 | 199,844 | 199,099 |
| 152 | 2499 | 8,140584 | -0,51626 | 2,486923 | 17,73029 | 31 | 199,844 | 199,099 |
| 153 | 2437 | 7,959905 | -0,55688 | 2,487847 | 17,3927 | 30 | 199,844 | 199,099 |
| 154 | 2373 | 7,788802 | -0,59381 | 2,486574 | 17,05268 | 29 | 199,844 | 199,099 |
| 155 | 2301 | 7,612908 | -0,62939 | 2,481693 | 16,67884 | 28 | 199,844 | 199,099 |
| 156 | 2238 | 7,475118 | -0,65341 | 2,474412 | 16,36014 | 27 | 199,844 | 199,099 |
| 157 | 2151 | 7,305633 | -0,67635 | 2,459565 | 15,9298 | 26 | 199,844 | 199,099 |
| 158 | 2046 | 7,116083 | -0,69667 | 2,432508 | 15,413 | 25 | 199,844 | 199,099 |

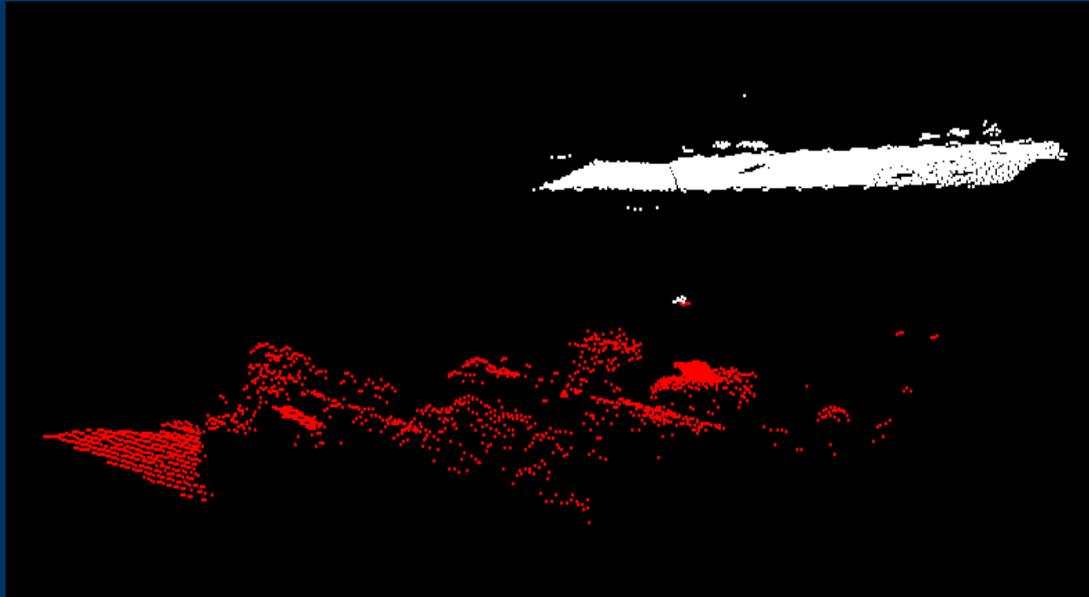
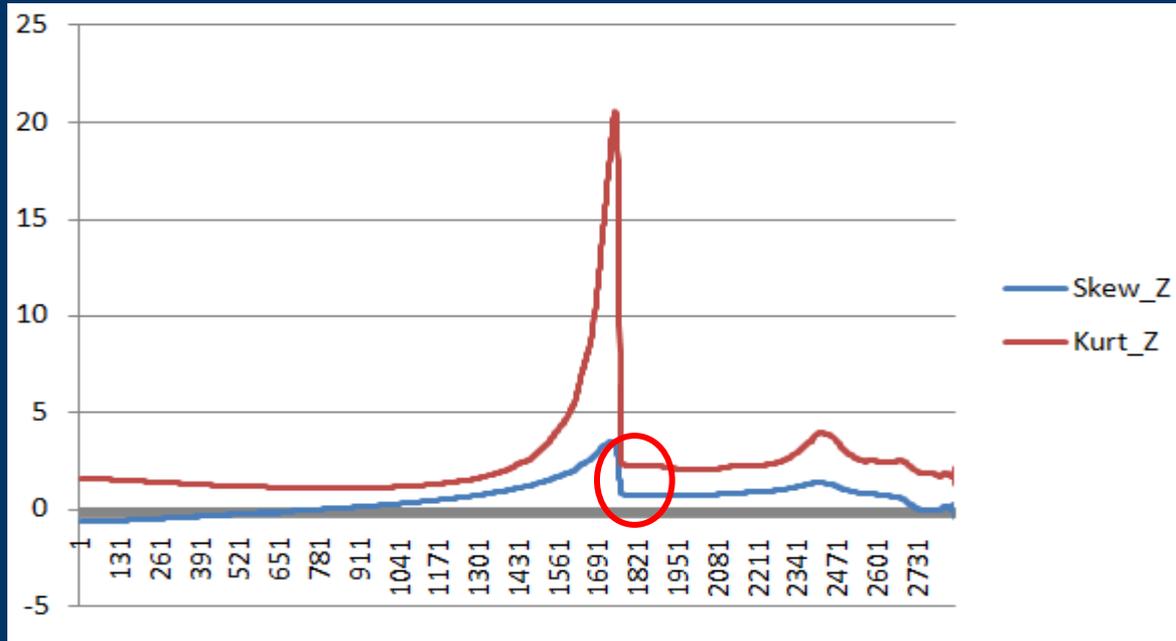


SECONDO ESPERIMENTO (4)

Valutazione punti bianchi

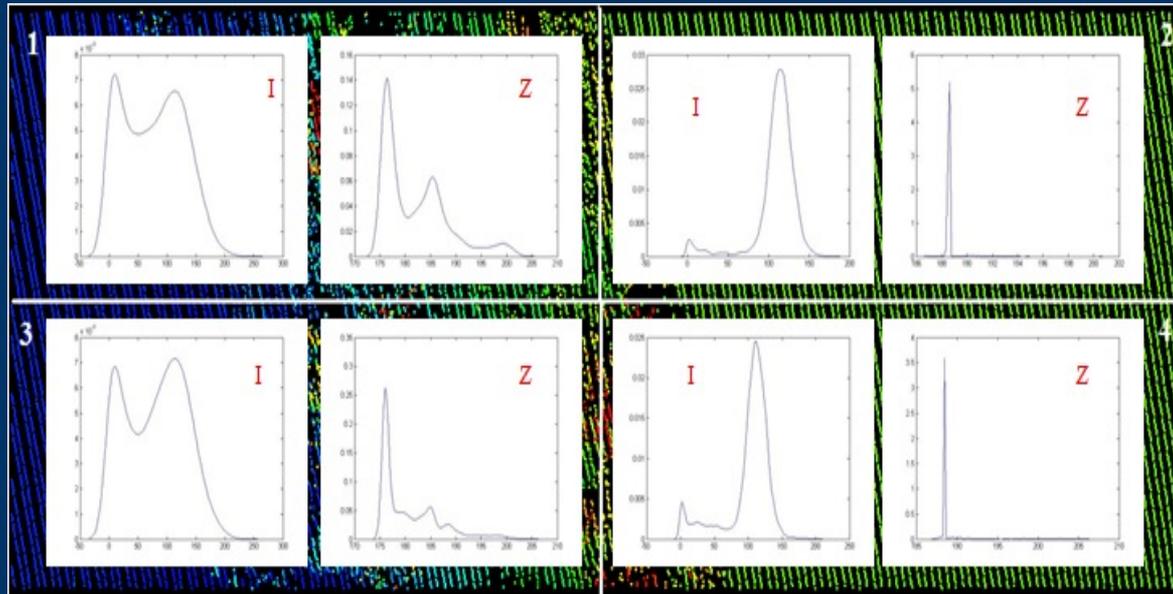
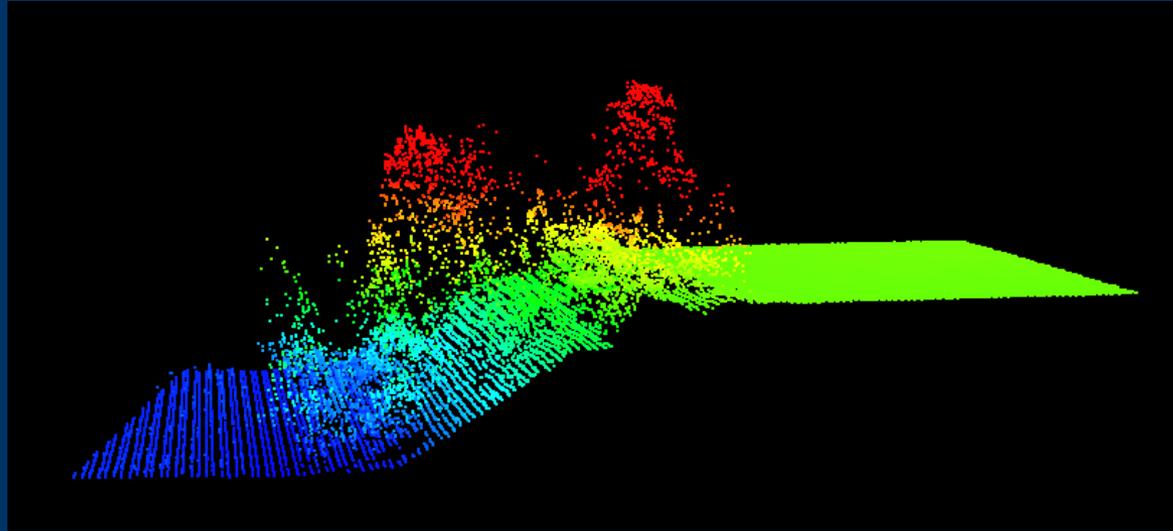


SECONDO ESPERIMENTO (5)

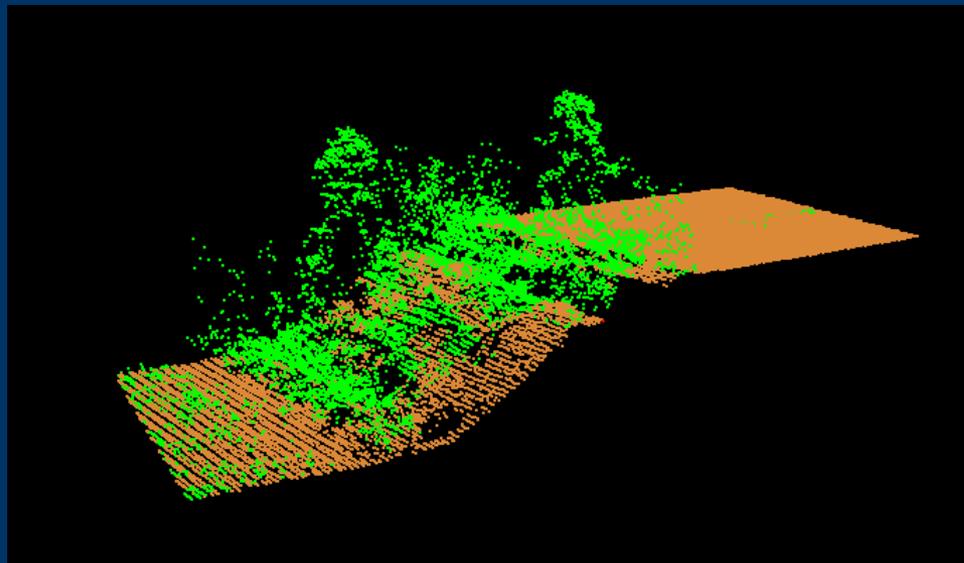
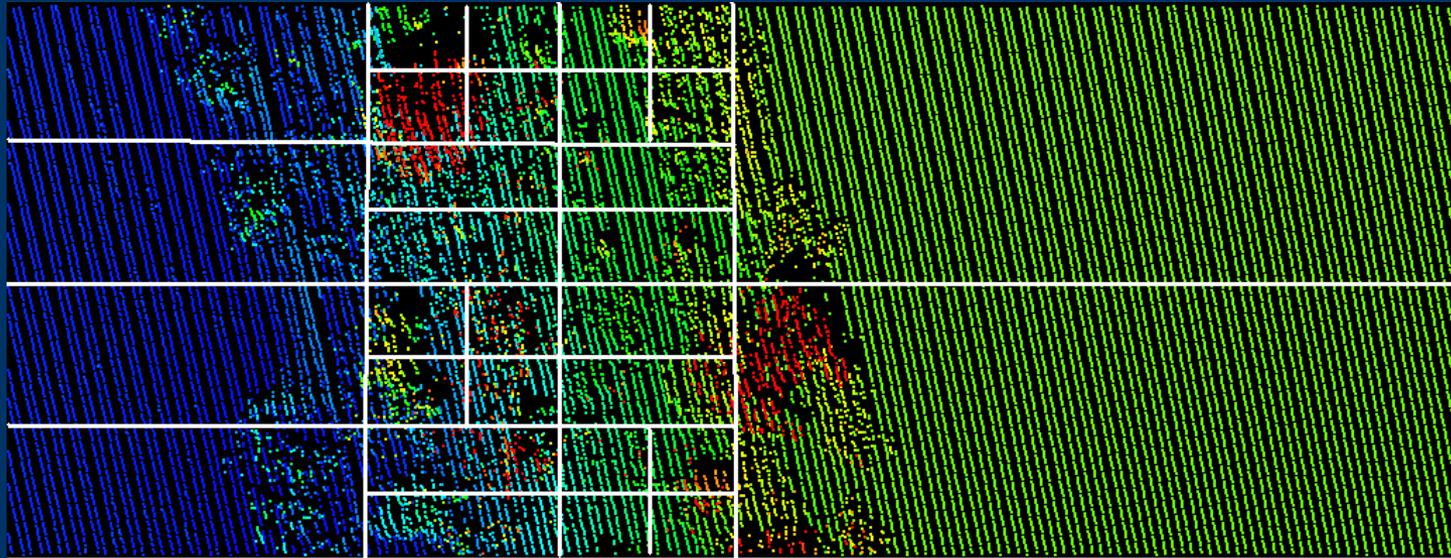


SUDDIVISIONE DELL'AREA (1)

Con questo esperimento si è testato il metodo di suddivisione dell'area in sottozone per poter applicare il metodo di classificazione in aree di difficile classificazione

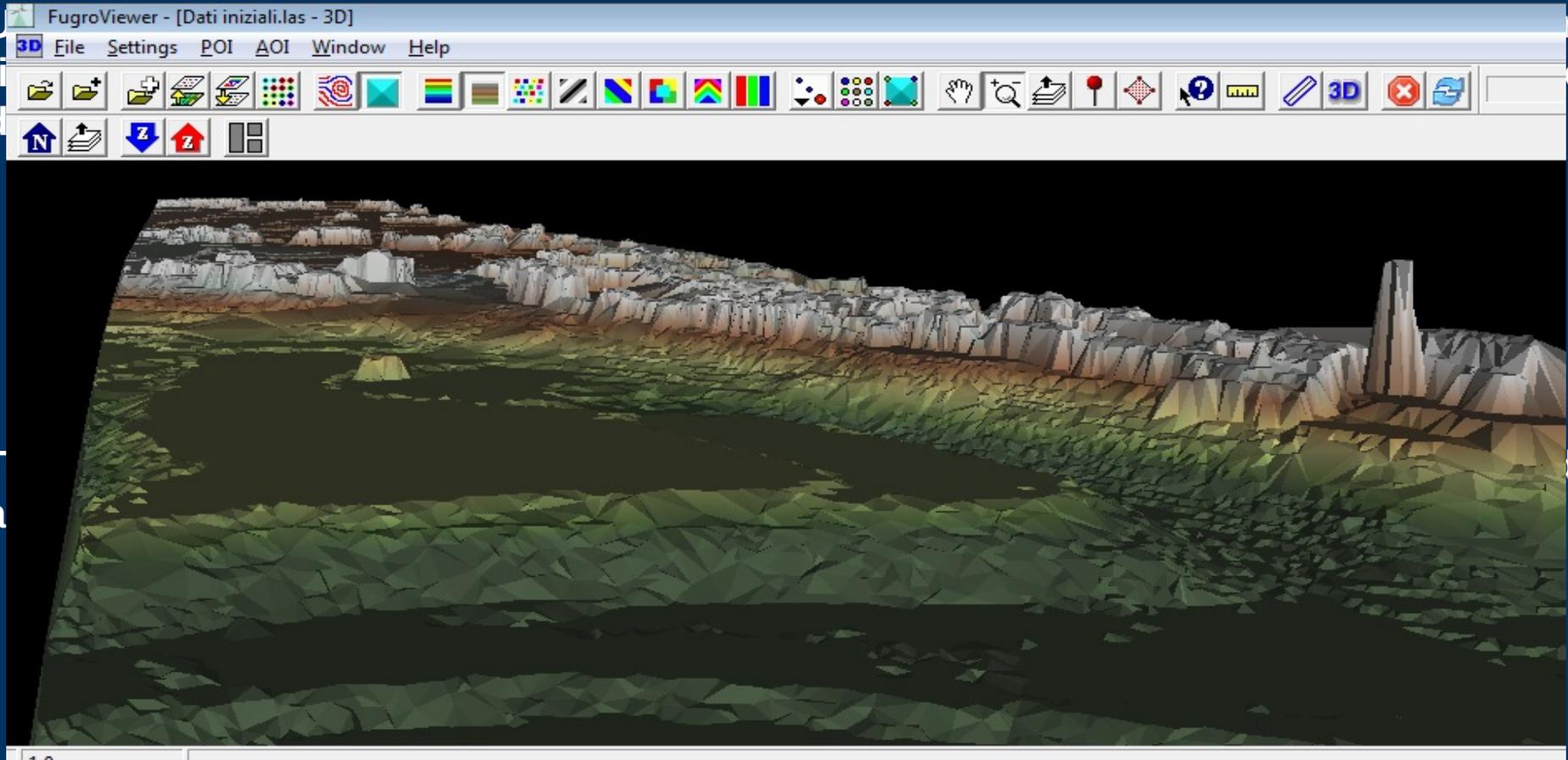


SUDDIVISIONE DELL'AREA (2)



Nell'area in esame (2805 mq) sono presenti 29131 punti di cui solo l'8,9% non è stato correttamente classificato

UTILIZZO DEL METODO DI CLASSIFICAZIONE CON PROGRAMMI OPENSOURCE



Creare applicativi per l'utilizzo del metodo in programmi OpenGIS (**Dielmo Open Lidar**) con la possibilità di visualizzazione una grossa mole di dati in ambiente 3D rende fruibile a tutti il metodo e ne consente una corretta visualizzazione

CONCLUSIONI

Lo studio svolto consente di:

- ∅ **Acquisire** la conoscenza degli elementi fondamentali per la classificazione dei dati laser mediante l'analisi degli indici di asimmetria e curtosi ampliando quanto già visto in letteratura con l'analisi della distribuzione dei punti e la suddivisione delle aree
- ∅ **Migliorare** la tecnica di classificazione di tipo statistico rendendola interattiva e sviluppando una metodologia per applicare sia il calcolo per l'intensità che per la geometria
- ∅ **Applicare** il metodo di classificazione attraverso l'utilizzo di programmi opensource

“siamo aperti a collaborazioni sull'argomento”

GRAZIE DELL'ATTENZIONE