

Studio della correlazione tra i parametri ambientali Umidità Relativa e Temperatura

Relazione di Laboratorio di Fisica 4 Mod.A
Anno Accademico 2020/2021

Gruppo 7

Cirillo Lorenzo - N94000701

Di Meo Paolo - N94000726

Favilla Leonardo - N94000698

Sommario

L'esperienza svolta mira a verificare la relazione funzionale che lega umidità relativa e temperatura in un sistema termodinamico ben definito. In questo elaborato, vengono presentate le varie fasi del procedimento, a partire dalla presa dei dati sperimentali, avvenuta da remoto con un sensore SHT75 ed una FPGA Nexys-4 tramite un programma di acquisizione scritto in linguaggio C, fino all'analisi degli stessi tramite programmi scritti in linguaggio C++ con l'utilizzo del software ROOT. Infine, vengono tratte le opportune conclusioni sulla base dei risultati dell'analisi dati effettuata.

Indice

1	Correlazione tra RH e T	2
2	Descrizione della strumentazione	4
2.1	Sensore SHT75	4
2.1.1	Misura di RH	4
2.1.2	Misura di T	5
2.1.3	Prestazioni del sensore	6
2.1.4	Comunicazione e trasmissione dati tra SHT75 e NEXYS-4	7
2.1.5	Conversione del dato	10
2.2	Scheda Nexys-4	10
3	Presa dati	12
3.1	Programma di acquisizione dati	12
3.1.1	Avvio dell'acquisizione	12
3.1.2	Struttura dei dati	13
3.1.3	Conversione dei dati	16
4	Analisi dati	17
4.1	Analisi preliminare	17
4.2	Suddivisione in sottocampioni	22
4.2.1	10 sottocampioni	23
4.2.2	20 sottocampioni	25
4.2.3	Test d'ipotesi	29
4.3	Sequenze temporali	38
	Appendice A: Tabella dei primi 100 eventi	48
	Appendice B: Esercizio di Machine Learning	51

1 Correlazione tra RH e T

Dato un sistema termodinamico definito da una pressione p ed una temperatura T costituito da vapore d'acqua in aria, definiamo *Relative Humidity* la quantità:

$$RH(\%) = 100 \frac{p(T)}{p_S(T)} \quad (1)$$

dove:

- i) $p_S(T)$ è la cosiddetta pressione di saturazione, ovvero il valore massimo di pressione p a quella data temperatura T prima che il vapore condensi in acqua;
- ii) $p(T)$ è la pressione a fissata temperatura T .

Se il sistema è costituito da un miscuglio di gas, vale la *Legge di Dalton*, secondo cui:

$$(p_1 + p_2 + \dots + p_n)V = (N_1 + N_2 + \dots + N_n)RT \quad (2)$$

dove abbiamo indicato con p_k ed N_k con $k = 1, \dots, n$ la pressione parziale e il numero di moli della specie k -esima presente nella miscela.

Osservando il diagramma delle fasi per l'acqua (Fig. 1), si può scrivere che:

$$\frac{RH(\%)}{100} = \frac{p(T)}{p_S(T)} = \frac{p_S(T_d)}{p_S(T)} \quad (3)$$

con T_d temperatura di rugiada, cioè la temperatura a cui bisogna portare il gas, a quella data pressione, affinché si instauri la fase liquida.

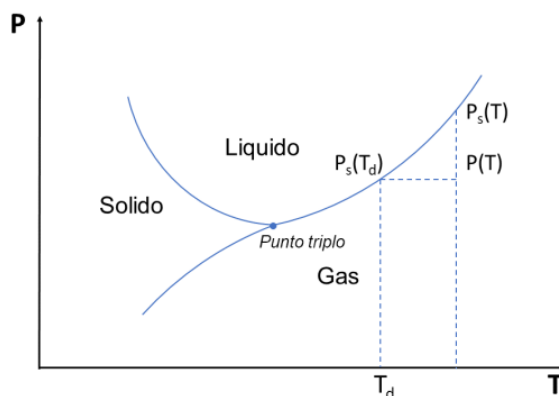


Figura 1: Diagramma della fasi per l'acqua.

Sappiamo che esiste una relazione, l'*Equazione di Clausius-Clapeyron*, che lega la pressione di saturazione p_S e la temperatura T :

$$\frac{dp_S}{dT} = \frac{L}{R_W} \frac{p_S}{T^2} \quad (4)$$

con:

- i) L = entalpia di vaporizzazione, ovvero l'energia da fornire ad 1gr di acqua per farla passare da fase liquida a gassosa. Essa varia tra $(2.26 - 2.50) \frac{J}{Kg}$ per una temperatura che varia tra $(0 - 100)^\circ C$;
- ii) R_W = costante del gas vapore d'acqua = $461.5 JK^{-1} kg^{-1}$.

Risolvendo tale equazione, si ottiene:

$$p_S(T) = kexp(-\frac{L}{R_W} \frac{1}{T}) \quad (5)$$

da cui:

$$\frac{RH(\%)}{100} = \frac{p(T)}{p_S(T)} = \frac{p_S(T_d)}{p_S(T)} = exp[-\frac{L}{R_W}(\frac{1}{T_d} - \frac{1}{T})] = exp[-\frac{L}{R_W}(\frac{T - T_d}{TT_d})] \quad (6)$$

Tuttavia, in genere si ha che $\frac{L}{R_W} \frac{T - T_d}{TT_d} \ll 1$, per cui si può sviluppare in serie di Taylor l'esponenziale e troncare al primo ordine, ottenendo la relazione:

$$\frac{RH(\%)}{100} = 1 - \frac{L}{R_W}(\frac{T - T_d}{TT_d}) \Rightarrow RH(\%) = 100 - (\frac{100L}{R_W} \frac{1}{TT_d})(T - T_d) \quad (7)$$

Assumiamo che possiamo scrivere:

$$RH(\%) = 100 - \beta(T - T_d) \quad \text{con } \beta \simeq \text{costante} \quad (8)$$

Calcolando β a $T = 293.15K$ si ha $\beta \simeq 5$, per cui ad una variazione di temperatura $\Delta T = 1K$ corrisponde una variazione di umidità relativa $\Delta RH(\%) = 5\%$.

Operando in un ambiente chiuso, la variazione della temperatura sarà dell'ordine del grado, a cui corrisponde, come visto, ad una variazione dell'ordine del percentuale per l'umidità relativa. Dunque, si richiede una risoluzione dello strumento che sia di:

$$O(10^{-2} \Delta T) \simeq 0.01^\circ C \text{ per la temperatura}$$

$$O(10^{-2} \Delta RH) \simeq 0.05\% \text{ per l'umidità relativa}$$

2 Descrizione della strumentazione

La strumentazione utilizzata per la presa dati consiste di:

- i) **Computer Portatile Acer**, con sistema operativo Ubuntu, cui si è fatto accesso con collegamento da remoto tramite VPN;
- ii) **Sensore SHT75 (Sensirion)** per la misurazione di umidità relativa e temperatura;
- iii) **Scheda NEXYS4 (Diligent)**, con Artix-7 FPGA già programmata, per gestire tutta la catena di acquisizione e processamento dei segnali.

2.1 Sensore SHT75

Le misurazioni di umidità relativa (RH) e temperatura (T) sono state fatte tramite l'utilizzo del sensore **SHT75** della Sensirion (Fig. 2).

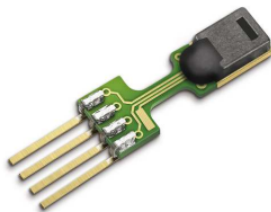


Figura 2: Sensore SHT75.

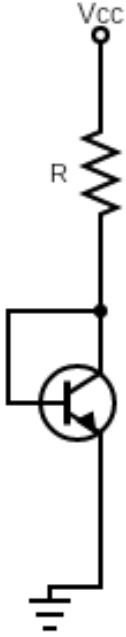
Il sensore (di dimensioni dell'ordine del mm) comprende gli elementi sensoriali propriamente detti che ed è in grado di processare segnali in forma compatta fornendo un output digitale. Analizziamo più nel dettaglio i processi di misurazione e di comunicazione con il sensore stesso.

2.1.1 Misura di RH

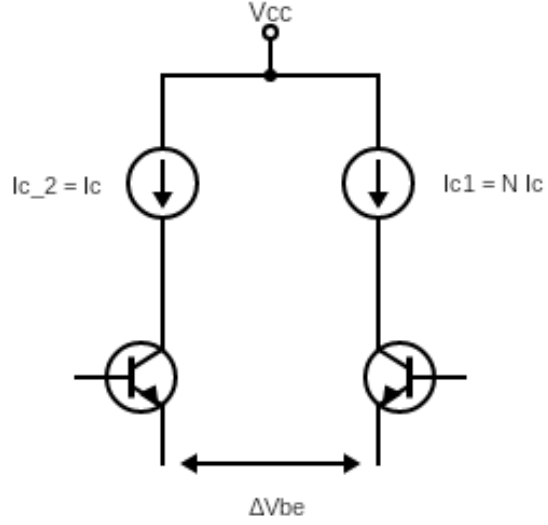
Un sensore capacitivo è l'elemento utilizzato per misurare l'umidità relativa. Sappiamo, infatti, che sussiste la relazione $Q = CV$ per un condensatore, per cui, fissata la carica Q depositata sullo stesso e misurata la differenza di potenziale V ai suoi capi, è possibile calcolare la capacità sua C , che, per un condensatore piano, sappiamo essere:

$$C = \frac{\epsilon_0 \epsilon_r (\text{vapore } H_2O) S}{d} \quad (9)$$

con S superficie delle lastre e d distanza tra di esse. Misurando C , si può dunque calcolare la costante dielettrica relativa $\epsilon_r (\text{vapore } H_2O)$ del vapore d'acqua, e da qui la densità del vapore nel volume considerato.



(a)



(b)

2.1.2 Misura di T

La misura di T viene, invece, effettuata tramite un sensore band-gap, un sensore a semiconduttore che sfrutta la dipendenza della differenza di tensione, misurata tra due transistor connessi a diodo, dalla temperatura. In particolare, alimentando un transistor connesso a diodo ($V_{BE} = V_{CE}$) con una corrente I_C costante si ottiene una tensione V_{BE} che varia linearmente con la temperatura T (Fig. 3a), infatti valgono:

$$V_{CC} = I_C R + V_{CE} \quad (10)$$

e

$$I_C = I_0 e^{\frac{V_{BE}}{V_T}} \quad \text{con} \quad V_T = \frac{k_B T}{q} \quad (11)$$

per cui:

$$V_{BE} = \ln\left(\frac{I_C}{I_0}\right) \frac{k_B T}{q} = C \frac{k_B T}{q} \quad \text{se} \quad I_C = \text{costante} \quad (12)$$

A questo punto, usando due transistor identici T1 e T2 ($I_{sat1} = I_{sat2}$) (Fig. 3b) e facendoli lavorare con un rapporto costante tra le correnti di collettore ($I_{C1} = N I_{C2} \equiv$

NI_C), si ha che:

$$\begin{aligned} V_{BE1} &= \ln\left(\frac{NI_C}{I_0}\right) \frac{k_B T}{q} = [\ln N + \ln\left(\frac{I_C}{I_0}\right)] \frac{k_B T}{q} = \\ &= \frac{k_B T}{q} \ln N + \underbrace{\ln\left(\frac{I_C}{I_0}\right) \frac{k_B T}{q}}_{V_{BE2}} \end{aligned} \quad (13)$$

ovvero:

$$\Delta V_{BE} = V_{BE1} - V_{BE2} = \frac{k_B T}{q} \ln N \quad (14)$$

2.1.3 Prestazioni del sensore

Entrambi i sensori sono accoppiati ad un convertitore ADC a 14 bit e ad un circuito di interfaccia seriale, che garantiscono una buona qualità del segnale ed un tempo di risposta rapido nel trasferimento dei dati verso l'esterno. Vediamo di seguito le varie caratteristiche del sensore e i valori corrispondenti:

- **Risoluzione:** Rappresenta l'incertezza statistica con cui viene misurata la grandezza in esame. Se la grandezza è esattamente costante, lo strumento la propone come distribuzione (tipicamente Gaussiana) centrata intorno al valore vero.
La misura di RH ha una risoluzione di $0.05\%RH$ (digitizzato a 12 bit), mentre per T è di $0.01^\circ C$ (digitizzato a 14 bit).
- **Accuratezza:** Rappresenta l'errore sistematico di cui è affetta la misura. Se la grandezza è esattamente costante, la distribuzione con cui viene proposta non sarà centrata sul valore vero, ed essa rappresenta la distanza tra valore medio della distribuzione e valore vero.
L'accuratezza è di $\pm 1.8\%RH$ per RH e di $\pm 0.3^\circ C$ per T.
- **Ripetibilità:** Se si effettuano due set di misurazioni della stessa grandezza nelle stesse condizioni, le due distribuzioni non saranno centrate sullo stesso valore. Rappresenta la distanza tra i due valori medi.
Essa vale $\pm 0.1\%RH$ per RH e di $\pm 0.1^\circ C$ per T.
- **Isteresi:** Differenza tra due valori della misura dello stesso valore vero dopo un ciclo temporale e di valori veri. Il valore misurato dipende dalla storia della grandezza.
Essa vale $\pm 1\%RH$ per RH e non si hanno valori per T.
- **Non linearità:** Differenza tra il valore atteso secondo relazione lineare e valore vero. Dipende dal punto di misura sulla scala di misura, tipicamente problematici bassi (entra in gioco la sensibilità dello strumento) e alti valori (il sistema satura e non distingue più tra due valori vicini).
Essa vale $\pm 3\%RH$ per RH e di $\pm 0^\circ C$ per T.
- **Tempo di risposta:** Tempo necessario affinché il sensore raggiunga il 63% di una variazione a gradino della grandezza misurata.
Essa vale 8s per RH e di 5-30s per T.

- **Intervallo di funzionamento:** Intervallo di valori della grandezza in cui la misura del sensore è fattibile senza danneggiare il dispositivo.
Essa vale $0 \div 100\%RH$ per RH e di $-40 \div 123.8^{\circ}C$ per T.
- **Deriva su tempi lunghi:** Variazione di risposta del sensore dopo un anno su un sistema termostato nelle stesse condizioni.
Essa è minore di $0.5\%RH/yr$ per RH e di $0.4^{\circ}C/yr$ per T.

2.1.4 Comunicazione e trasmissione dati tra SHT75 e NEXYS-4

Il sistema di acquisizione e trasmissione dati è costituito da un collegamento, tramite i suoi 4 pin, tra il sensore SHT75 e un Microcontrollore (Fig. 4), che è costituito da una FPGA (*Field Programmable Gate Array*), ovvero un dispositivo elettronico programmabile formato da un circuito integrato in grado di gestire il trasferimento dati. In questo caso è stata usata una scheda NEXYS-4 come microcontrollore.

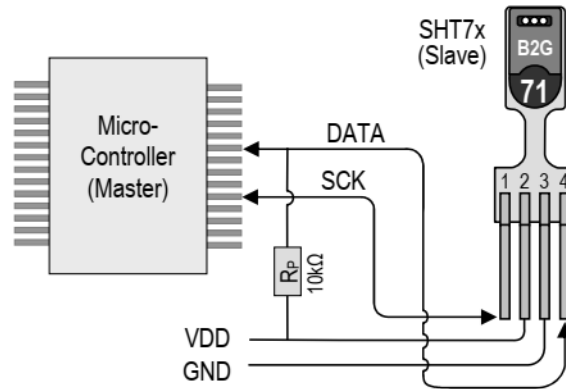


Figura 4: Circuito per trasmissione dati.

I pin del sensore SHT75 sono:

- Pin 1) Serial Clock (SCK): viene usato per sincronizzare la comunicazione tra microcontrollore e SHT75;
- Pin 2) V_{DD} : viene fornita la tensione di alimentazione per un valore compreso tra 2.4V e 5.5V (valore consigliato per V_{DD} è 3.3V);
- Pin 3) Linea Dati Seriale (DATA): un dispositivo a tre stati viene utilizzato per trasferire i dati in entrata ed uscita dal sensore.
- Pin 4) Ground (GND).

Per inviare un comando al sensore, la linea DATA è validata solo sul fronte di salita del SCK e deve rimanere stabile mentre SCK è alto. Sul fronte di discesa di SCK il valore DATA può essere modificato (Fig. 5). Per sicurezza, la linea DATA viene estesa di un tempo T_{SU} prima del fronte di salita e T_{HO} dopo il fronte di discesa di SCK. Per leggere

dati dal sensore, DATA è valido a partire da un tempo T_V dopo che SCK è andato basso e resta valido fino al prossimo fronte di discesa del SCK.

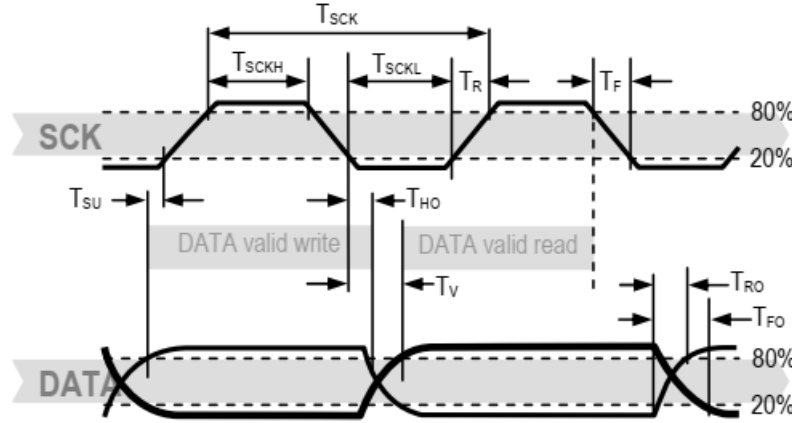


Figura 5: *Timing Diagram* delle linee DATA e SCK. La linea DATA spesso è controllata dal sensore, mentre quella sottile dal microcontrollore.

Da notare che lo stato normale della linea DATA è quello alto, per cui viene inserita una resistenza R_P di pull-up tra le linee DATA e di alimentazione V_{DD} per riportare in alto la linea DATA quando necessario. Inoltre, vediamo che il microcontrollore, prima di inviare un comando al sensore, deve avviare una sequenza di inizio trasmissione (Fig. 6) che consiste nell'abbassamento della linea DATA mentre il SCK è alto, seguito da un impulso basso su SCK mentre DATA è ancora bassa e poi di nuovo livello alto di DATA mentre SCK è alto.

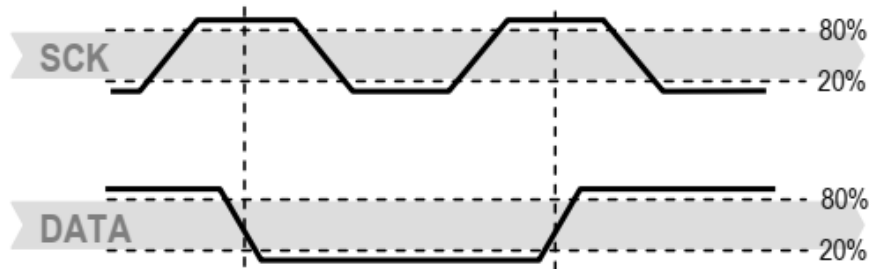


Figura 6: Sequenza di *Inizio Trasmissione*.

Il comando successivo consiste in 3 bit di indirizzo (solo **000** è supportato dato che il sensore ha un solo indirizzo) e 5 di comando. Per avviare la misurazione, i 5 bit di comando sono:

- Per la temperatura: **00011**;
- Per l'umidità relativa: **00101**.

Al termine della trasmissione del comando dal microcontrollore al sensore, quest'ultimo comunica la corretta ricezione abbassando la linea DATA dopo il fronte di discesa dell'ottavo colpo di SCK (bit **ACK** - *Acknowledge*) e rilasciandola dopo il nono impulso di clock (Fig. 7). In seguito a ciò, la linea DATA rimane inattiva (quindi alta) durante il tempo di svolgimento del comando da parte del sensore (Questo richiede un massimo di 20/80/320 ms per una misura a 8/12/14 bit). Per segnalare il completamento di una misura, il sensore abbassa la linea DATA ed entra in modalità inattiva. Il microcontrollore, alla ricezione di questo segnale di **Data Ready**, riabilita SCK per leggere i dati, che vengono memorizzati sull'SHT75 fino alla lettura.

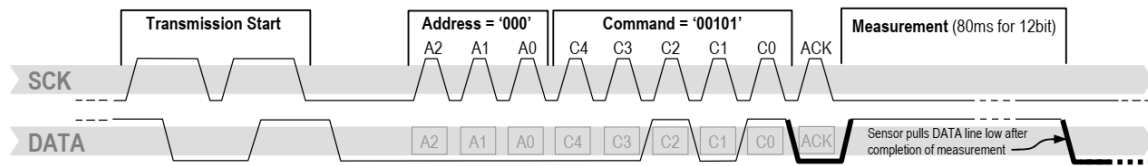


Figura 7: Sequenza di comando per la misurazione di RH. La linea DATA sottile è pilotata dal microcontrollore, mentre quella spessa dal sensore.

A questo punto, il sensore invia 2 bytes di dati di misurazione ed 1 byte di *CRC Checksum* (opzionale) al microcontrollore, il quale comunica l'avvenuta ricezione abbassando la linea DATA al termine di ogni byte. La sequenza è ordinata a partire dal *Most Significant Bit* (MSb) al *Least Significant Bit* (LSb). Poichè la codifica è fatta a 12 (14) bit per RH (T), i dati sono giustificati a destra, ossia i primi 4 (2) bit sono inutili (**Idle bits**) e vengono posti a 0; a partire dal quinto bit, si ha la comunicazione del dato secondo il protocollo mostrato.

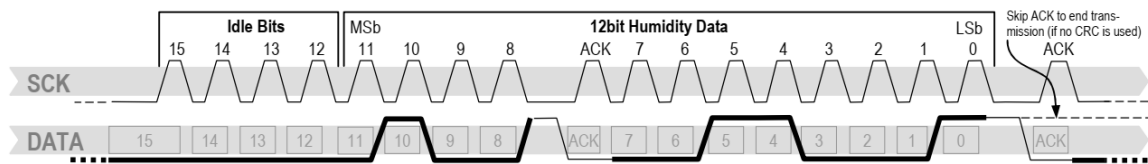


Figura 8: Invio dati di misura di RH da SHT75 alla NEXYS-4: si vedano i primi 4 idle bits, seguiti dai 12 bits di codifica del dato dal MSb al LSb e 2 bits di ACK.

La comunicazione termina dopo 1 bit di ACK relativo al byte di CRC (Fig. 9). Se tale ciclo CRC non viene utilizzato, il controllore può terminare la comunicazione dopo la trasmissione del LSb mantenendo ACK alto.

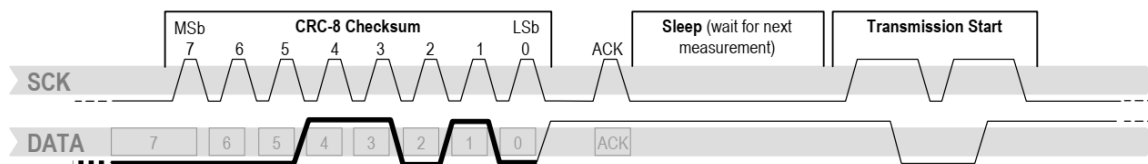


Figura 9: Sequenza CRC (se utilizzata).

Il dispositivo torna automaticamente in modalità riposo dopo la misurazione fino alla trasmissione successiva.

2.1.5 Conversione del dato

Il sensore restituisce il risultato delle sue misure sotto forma di un intero nell'intervallo $[0, 4095]$ per l'umidità e $[0, 16383]$ per la temperatura. Per compensare la non linearità del sensore di umidità, il valore intero RH_{raw} in ADC count va convertito secondo la formula:

$$RH_{lin}(\%) = a + bRH_{raw} + c(RH_{raw}^2) \quad (15)$$

$$\text{con } a = -2.0468, \quad b = 0.0367, \quad c = -1.5955 \times 10^{-6}$$

Va inoltre aggiunta una correzione proporzionale alla differenza tra la temperatura di misura e $25^\circ C$, dunque:

$$RH_{true}(\%) = (T - 25) \cdot (t_1 + t_2 \cdot RH_{raw}) + RH_{lin} \quad (16)$$

Per la temperatura vale invece una relazione lineare:

$$T(^{\circ}C) = a + bT_{raw} \quad (17)$$

$$\text{con } a = -39.7, \quad b = 0.01$$

2.2 Scheda Nexys-4

Una *Field Programmable Gate Array* (FPGA) è un dispositivo logico programmabile, ovvero un dispositivo elettronico formato da un circuito integrato le cui funzionalità logiche sono programmabili e modificabili tramite opportuni linguaggi di programmazione.

La scheda Nexys-4 è un circuito digitale completo su cui è installata l'FPGA Artix-7. Essa presenta una porta USB, predisposta alla comunicazione con pc, una memoria RAM (4860 kbits) e 4 connettori per l'entrata di periferiche digitali (tra cui il sensore SHT75) (Fig. 10).

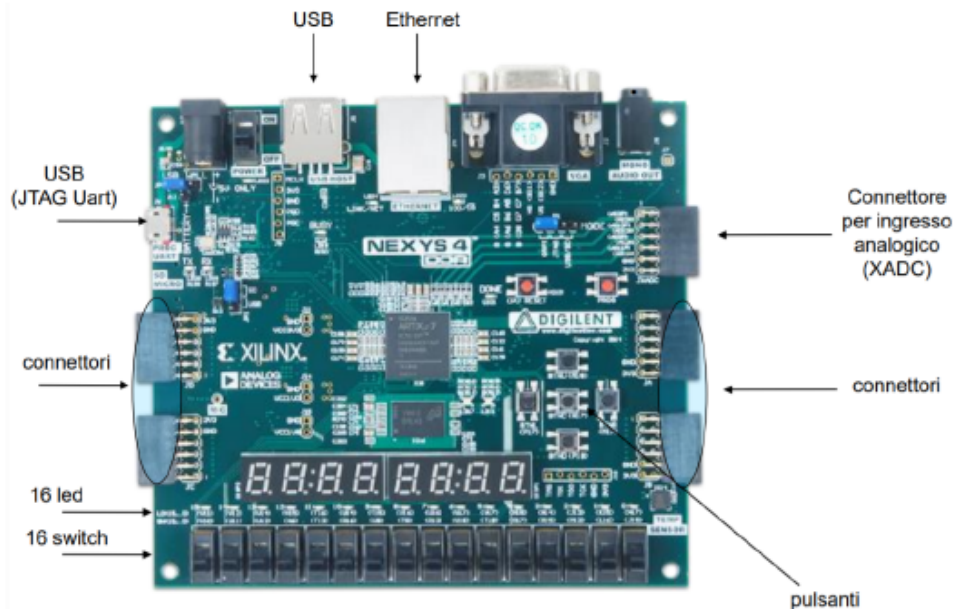


Figura 10: NEXYS-4.

La Nexys-4 funziona tramite il firmware caricato sulla FPGA, che è programmata in modo tale da richiedere al sensore l'umidità relativa e la temperatura in sequenza, con una frequenza di circa 2.5 Hz, quindi ogni 2 secondi si hanno 5 misure della coppia umidità relativa e temperatura. La scheda gestisce i sensori, fornendo l'alimentazione e il riferimento GND e impacchetta i dati conservandoli in una memoria. I dati vengono infine trasferiti al laptop secondo il **protocollo RS232**. Il protocollo RS232 (Recommended Standard 232) consiste in un formato particolare di trasmissione di dati, ad una fissata frequenza di trasferimento, detta *baud rate* (4800, 9600, 38400 o 115200 bps), strutturato come segue:

- 1 bit di START;
- 8 bits di DATI;
- 1 bit di PARITÀ;
- 1 bit di STOP.

Il trasferimento dei dati dalla scheda al PC avviene in pacchetti di N bytes. Non avendo un controllo intrinseco degli errori (solo il bit di parità), si può realizzare un controllo personalizzato, aggiungendo a ciascun pacchetto una parola di testa e/o di coda ed eventualmente altre parole di controllo. Nell'invio al PC i dati raccolti dal sensore, la Nexys-4 utilizza una parola di controllo costituita da 2 bytes. Tale parola di controllo è la stringa: "1010'1010" "1010'1010". In seguito a questa, vengono trasmessi 4 bytes di misura (2 per RH_{raw} e 2 per T_{raw}).

3 Presa dati

La presa dati è avvenuta da remoto, tramite accesso VPN al PC 2 del laboratorio dell'università, il giorno 03/12/2020 dalle ore 15:00 alle ore 18:00. L'ambiente dove è stata svolta la misura era chiuso e sottoposto a climatizzazione. Si sono raccolti quindi 28780 coppie di dati relativi a umidità relativa e temperatura.

3.1 Programma di acquisizione dati

3.1.1 Avvio dell'acquisizione

L'acquisizione dei dati è avvenuta tramite un programma scritto in linguaggio C, facendo uso di librerie specifiche per comunicare con la Nexys-4, trasferire i dati sotto forma di pacchetti di bytes e scriverli su dei file di testo. Di fondamentale importanza è la libreria

```
19 #include "rs232.h"
```

che contiene tutte le funzioni che gestiscono la comunicazione con la scheda. Inoltre la risoluzione necessaria a campionare le misure impone l'utilizzo della funzione **gettimeofday**, che restituisce i microsecondi (μs) trascorsi dalla mezzanotte del giorno corrente. Tale funzione necessita delle librerie

```
10 #include <sys/time.h>
11 #include <time.h>
```

Nel main viene chiesto preliminarmente in input il tempo di acquisizione (in ore) e vengono creati, oltre alle variabili necessarie, dei file di testo, tra cui uno che contenga i dati grezzi provenienti dalla scheda ed un secondo per i valori delle misure convertiti secondo le formule (15), (16) e (17).

```
25 int main(int argc, char *argv[])
26 {
27
28     struct tm *gmp;
29     struct tm gm;
30     time_t t, t0;
31     int ty, tmon, tday, thour, tmin, tsec, time_acq_h_MAX;
32     float time_acq_h;
33     int i, n, nloc, InitFlag, StartFlag, nhit, hit, trg, cport_nr=17, bdrate=115200;
34     char pacchetto[9] = "pacchetto";
35     struct timeval current_time_0;
36     struct timeval current_time;
37     int j;
38
39
40     int cnt=0;
41     FILE *file;
42     FILE *currN; //file that stores current run name to be used by external programs
43     FILE *pacchetti;
44     double val_temp, val_hum_corrected;
45     unsigned char buf[4096], sht75_nblab02_frame[4];
46     unsigned int val_temp_int, val_hum_int;
```

```

47 char NameF[100];
48
49 char mode[]={'8','N','1',0};

```

Alla riga 33 vengono definiti due interi: **cport_nr**, ossia il numero della porta di comunicazione alla quale è connessa la scheda, e **bdrate**, ovvero la baud rate del trasferimento. Alla linea 49 viene definito un array di char che contiene tutti i parametri per la trasmissione dei dati: rispettivamente si intende un trasferimento a pacchetti di 8 bits, senza bit di parità, uno di chiusura e senza controllo del flusso. Nella riga 45 si definisce l'array di char **buf** che ospiterà i dati inviati dalla Nexys-4. Il motivo per il quale si è scelta la tipologia *char* è che i dati sono codificati in esadecimale.

In seguito a ciò, il programma acquisisce l'orario corrente ed invoca una funzione della libreria **rs232.h** che permette di avviare la comunicazione con la scheda.

```

94 if(RS232_OpenComport(cport_nr, bdrate, mode))
95 {
96     printf("Can not open comport\n");
97     return(0);
98 }

```

La funzione richiede come input i parametri introdotti in precedenza e restituisce il valore 1 in caso di errore.

A questo punto ha inizio un ciclo **while** che si interrompe una volta superato il tempo massimo di acquisizione. In questo ciclo si richiama periodicamente la funzione **RS232_PollComport**, che svolge il compito di interrogare la Nexys-4 e prelevare tutti i dati presenti nella sua memoria al momento della richiesta. La funzione ammette come argomenti il numero della porta seriale, l'array in cui depositare i pacchetti acquisiti e il numero massimo di byte che possono essere immagazzinati e restituisce un intero **n** che corrisponde al numero di bytes ricevuti nel buffer.

```

107 while(1)
108 {
109
110     n = RS232_PollComport(cport_nr, buf, 4095);

```

```

124 if (time_acq_h > time_acq_h_MAX * 3600)
125 {
126     printf(" time_duration RUN in hours > %d \n", time_acq_h_MAX);
127     break;
128 }

```

3.1.2 Struttura dei dati

La trasmissione delle informazioni avviene secondo una struttura a pacchetti ben precisa definita nei *char mode*. I dati viaggiano nella seguente sequenza di bytes:

1. 2 bytes della parola di controllo "AA AA"
2. 2 bytes di lettura dell'umidità relativa

3. 2 bytes di lettura della temperatura

Il programma deve quindi ricercare la parola di controllo che introduce ogni pacchetto di dati così da poter acquisire i 4 bytes successivi, ossia i valori delle misure a cui siamo interessati.

```
132 for(i=0; i < n; i++)
133 {
134     if (i>0 && buf[i]==0xAA && buf[i-1]==0xAA)
135     {
136
137         StartFlag=1;
138         nloc=0;
139         if (InitFlag==0)
140         {
141             InitFlag=1;
142         }
143     }
```

In questo modo l'indicatore **InitFlag**, precedentemente posto a 0, assume il valore 1 solo nel caso in cui si sia identificata la parola di controllo. In tal caso può avere inizio la procedura di scrittura delle misure sui file di testo che serviranno per l'analisi.

```
152 if (InitFlag==1)
153 {
154     sht75_nblab02_frame[nloc]=buf[i];
155     if (nloc==1)
156     {
157         printf("pacchetto %d\n", trg);
158         fprintf(pacchetti, "pacchetto %d\n", trg);
159
160         printf("seconds : %f\n", (((current_time.tv_sec * 1000000 +
161         current_time.tv_usec) - (current_time_0.tv_sec * 1000000
162         + current_time_0.tv_usec))/1000000.));
163         fprintf(pacchetti, "seconds : %f\n", (((current_time.tv_sec * 1000000
164         + current_time.tv_usec) - (current_time_0.tv_sec * 1000000 +
165         current_time_0.tv_usec))/1000000.));
166
167         printf("AA AA\n");
168         fprintf(pacchetti, "AA AA\n");
169
170         printf("%x %x\n", sht75_nblab02_frame[0], sht75_nblab02_frame[1]);
171         fprintf(pacchetti, "%x %x\n", sht75_nblab02_frame[0], sht75_nblab02_frame[1]);
172
173         val_hum_int=(sht75_nblab02_frame[0]<<8)|(sht75_nblab02_frame[1]);
174
175         val_hum_corrected = decodeHumidity(val_hum_int, 25, val_temp);
176
177         fprintf(file,"%d %d %d %d %.2f %f", trg, gm.tm_year+1900, gm.tm_mon+1,
178         gm.tm_mday, 3600*gm.tm_hour + 60*gm.tm_min + (float)gm.tm_sec,
179         (((current_time.tv_sec * 1000000 + current_time.tv_usec)
180         - (current_time_0.tv_sec * 1000000 + current_time_0.tv_usec))/1000000.));
181         fprintf(file, " %d %.2f ", val_hum_int, val_hum_corrected);
182     }
```

```

183         trg++;
184         hit++;
185     }
186     else if(nloc==3)
187     {
188         printf("%x %x\n", sht75_nblab02_frame[2], sht75_nblab02_frame[3]);
189         fprintf(pacchetti, "%x %x\n", sht75_nblab02_frame[2], sht75_nblab02_frame[3]);
190
191         val_temp_int=(sht75_nblab02_frame[2]<<8)|(sht75_nblab02_frame[3]);
192         val_temp=decodeTemperature(val_temp_int);
193
194         fprintf(file,"%d %.2f\n", val_temp_int, val_temp);
195     }
196
197     nloc++;

```

L'intero **nloc** aumenta da 0 a 5 così da riempire l'array **sht75_nblab02_frame** con tutti i dati del singolo pacchetto. Questi vengono stampati a video e scritti sul file di testo **pacchetti** insieme al tempo trascorso dall'inizio dell'acquisizione (righe da 154 a 171). La variabile **nloc** serve anche ad identificare i **MSB** dell'umidità (se **nloc**=1) e della temperatura (se **nloc**=3). Infine, il contatore **trg** numera i pacchetti. Il risultato di questa prima parte sarà una struttura del tipo

```

                pacchetto 62
seconds :   23.625285
        AA AA
        05 3f
        19 15

```

Nelle righe 173 e 191 vi sono le funzioni per convertire i valori di umidità relativa e temperatura da esadecimale a decimale. Gli interi ottenuti costituiscono l'input per le funzioni che li convertono nei rispettivi valori in percentuale e gradi centigradi.* I valori interi e quelli convertiti delle misure vengono quindi scritti in un secondo file di testo (chiamato **file**).

Siccome la comunicazione tra il sensore SHT75 e la scheda Nexys-4 avviene ad una frequenza di 2.5 Hz , la richiesta di trasferimento dei dati dalla scheda al PC deve avvenire ogni 400 ms . Tramite la funzione **usleep** il programma sospenderà la sua attività per il tempo impostato, prima di interrogare nuovamente la scheda.

```

234 #ifdef _WIN32
235     Sleep(400);
236 #else
237     usleep(400000);
238 #endif
239 }
240 fclose (pacchetti);
241 fclose (file);
242
243 return(0);
244 }

```

*Queste funzioni verranno espone nel prossimo paragrafo.

3.1.3 Conversione dei dati

Il valori di umidità relativa e temperatura ottenuti dal sensore sono interi tra 0 e $2n$, dove n è il numero di bit utilizzati dall'ADC counter. Per convertire quindi il dato *raw* in un dato fisicamente interpretabile, si è fatto riferimento alle informazioni contenute nel *datasheet* del sensore SHT75 per le formule di conversione e per i rispettivi coefficienti (eq. 15, 16 e 17):

```
248 double decodeTemperature(unsigned int rbuf) {
249     double d1, d2, rd_val;
250     d1=-39.6;
251     d2=0.01;
252     rd_val=(double)rbuf + 0.;
253     return d1+d2*rd_val ;
254 }
255
256 double decodeHumidity(unsigned int rbuf, double temperature_ref, double val_temp){
257     double c1, c2, c3, rd_val, hum_val, hum_val_corrected;
258     c1 = -2.0468;
259     c2 = 0.0367;
260     c3 = -1.5955e-6;
261     rd_val = (double)rbuf;
262     hum_val = c1+c2*rd_val+c3*(rd_val)*(rd_val);
263     hum_val_corrected = (val_temp - temperature_ref) * (0.01 + 0.00008 * rd_val) + hum_val;
264     //return hum_val;
265     return hum_val_corrected;
266 }
```

4 Analisi dati

Come anticipato nella prima parte, la relazione che si vuole studiare (eq. 8) è del tipo

$$RH = a + bT \quad (18)$$

Si intendono quindi stimare i parametri a , b ed il coefficiente di correlazione ρ tra le due grandezze. Nel corso delle 3 ore della presa dati l'umidità relativa ha avuto una variazione massima dell' 1.74%, mentre l'escursione termica è stata di circa $0.47^\circ C$.

4.1 Analisi preliminare

Una prima verifica della correlazione attesa si ottiene graficando l'andamento temporale delle due grandezze misurate (Fig.11 e12). Come ci si aspettava, le due grandezze mostrano un andamento speculare, suggerendo di conseguenza una correlazione negativa. Ciò risulta ancora più evidente nello scatter plot di RH in funzione di T (Fig. 13). Le coppie di punti sperimentali si dispongono abbastanza uniformemente attorno ad una retta con pendenza negativa, senza particolari irregolarità. Un'ulteriore indagine qualitativa sulle misure si effettua calcolando le medie temporali di RH e T ogni 200 e 500 punti. Considerando che in intervalli di tempo brevi ci si aspetta una variazione delle grandezze ragionevolmente piccola, il plot delle loro medie in funzione del tempo fornisce una descrizione aggiuntiva dell'evoluzione temporale del sistema (Fig. 18 e 19). Si individua dunque una regione all'interno della quale le due grandezze mantengono una separazione pressoché speculare, senza variazioni eccessivamente repentine, ed una regione, concentrata verso gli ultimi minuti della presa dati, nella quale la separazione si attenua forse a causa di una perturbazione dell'ambiente. In generale, le medie sembrano riprodurre ragionevolmente l'andamento del set intero di dati.

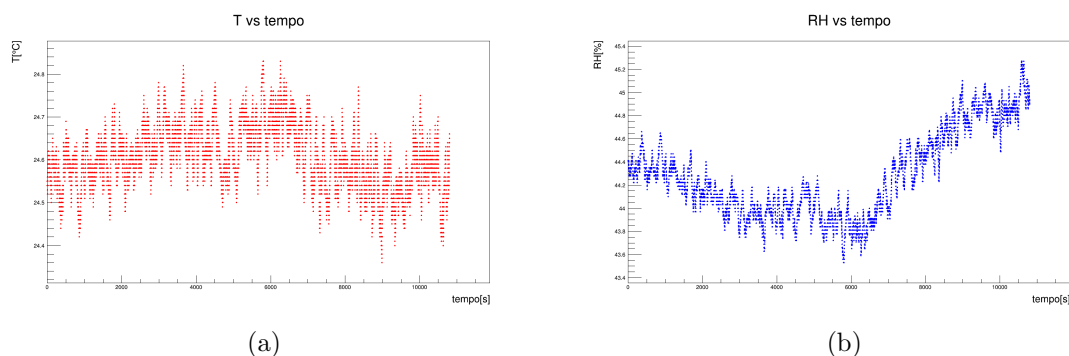


Figura 11: Andamento nel tempo della **temperatura** (a) e dell'**umidità relativa** (b).

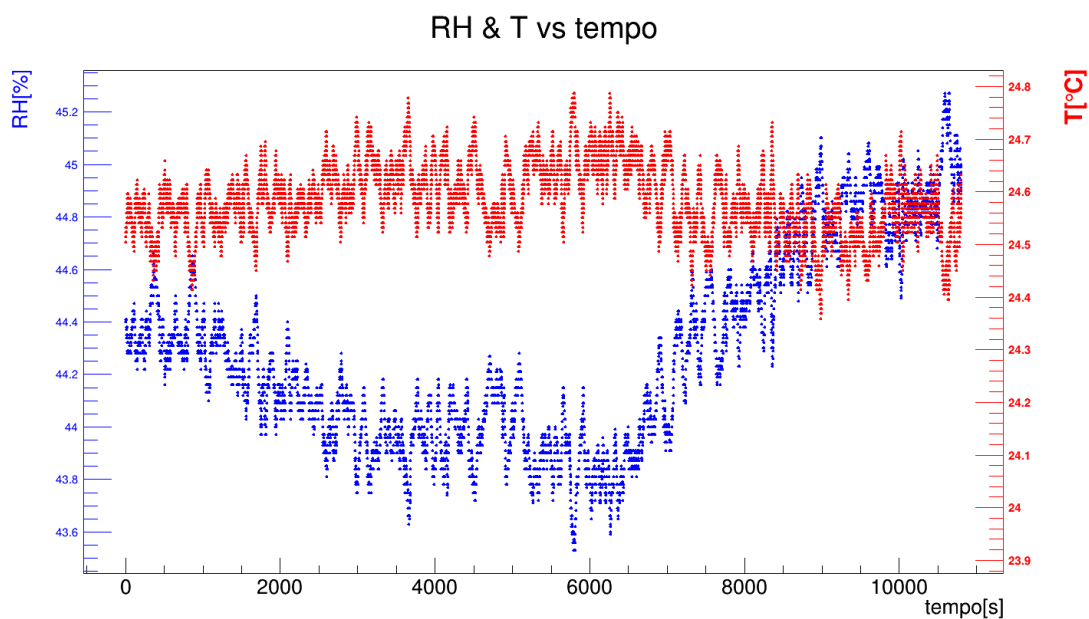


Figura 12: Andamento nel tempo della **temperatura** e dell'**umidità relativa** a confronto.

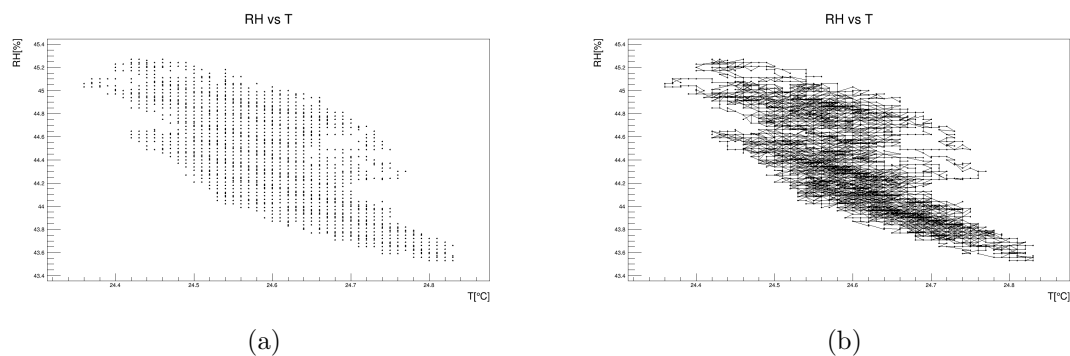


Figura 13: Scatter plot di RH e T con punti separati (a) e congiunti (b).

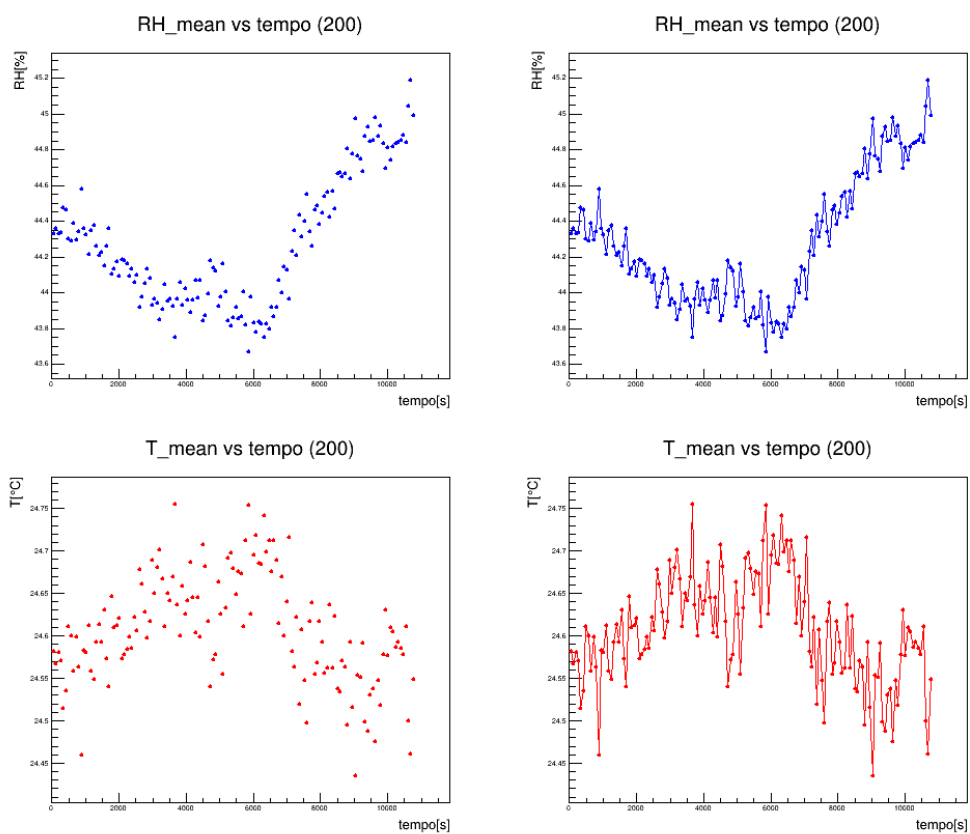


Figura 14: Andamento nel tempo dei valori medi su 200 punti della **temperatura** e dell'**umidità relativa**.

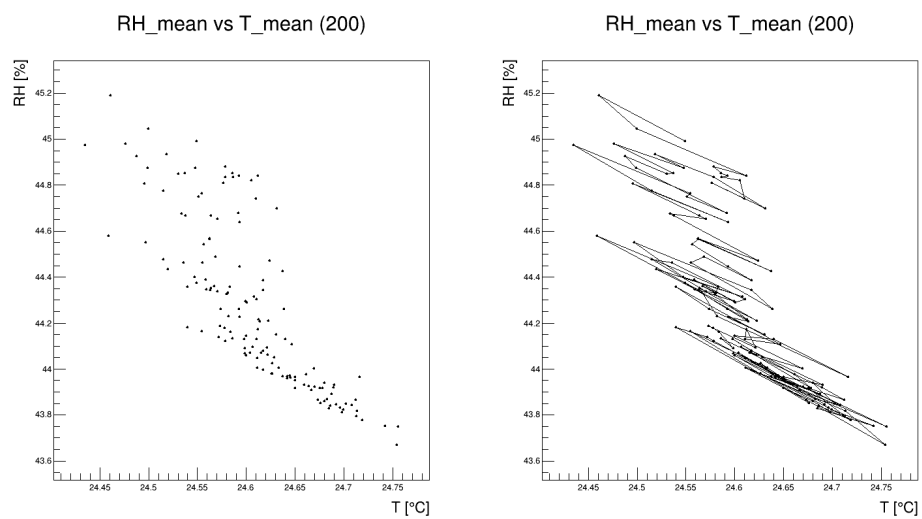


Figura 15: Scatter plot delle medie su 200 punti.

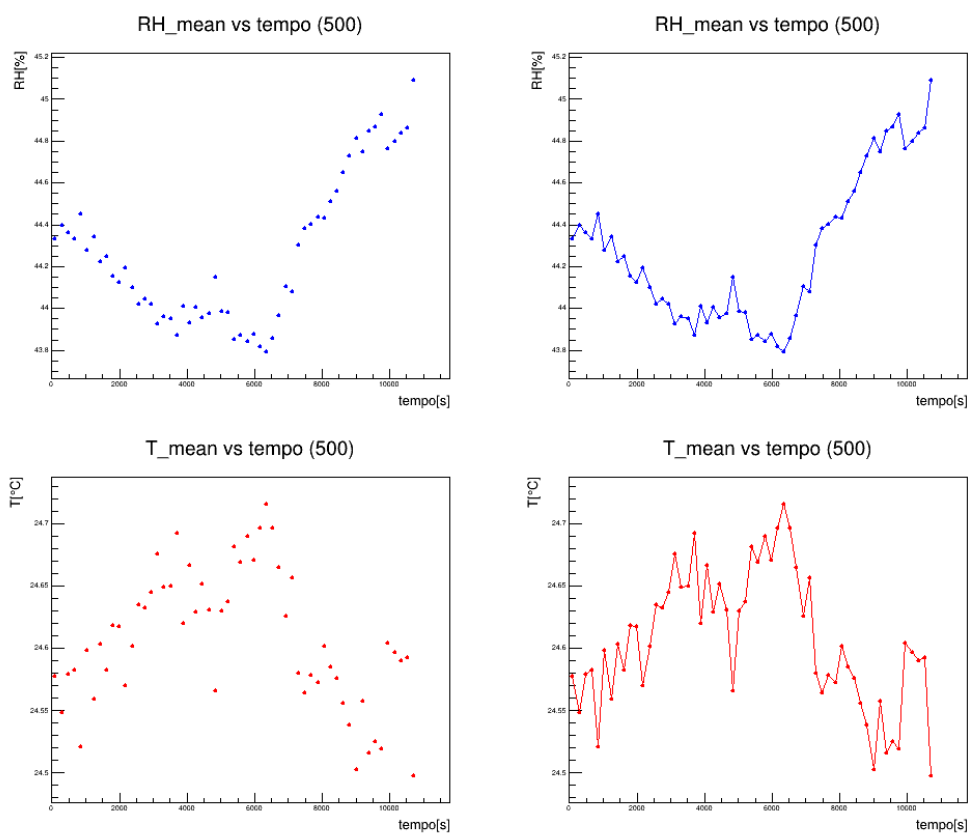


Figura 16: Andamento nel tempo dei valori medi su 500 punti della **temperatura** e dell'**umidità relativa**.

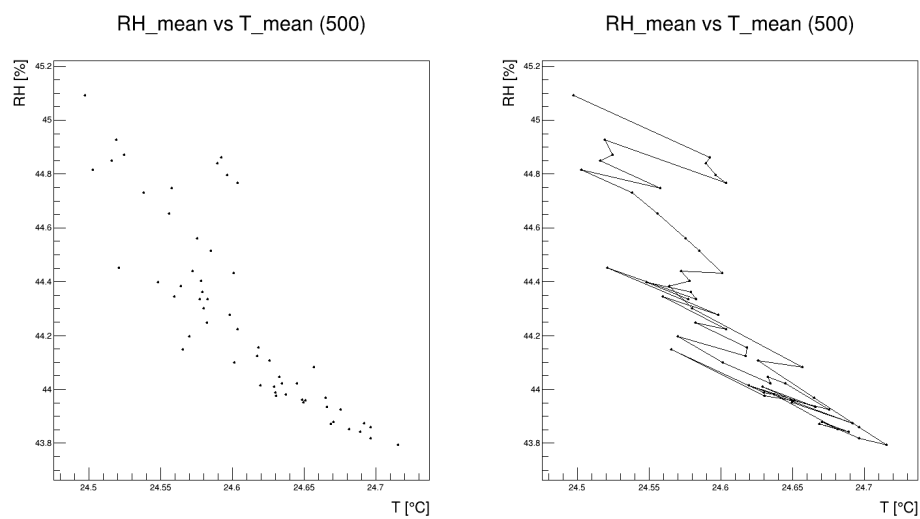


Figura 17: Scatter plot delle medie su 500 punti.

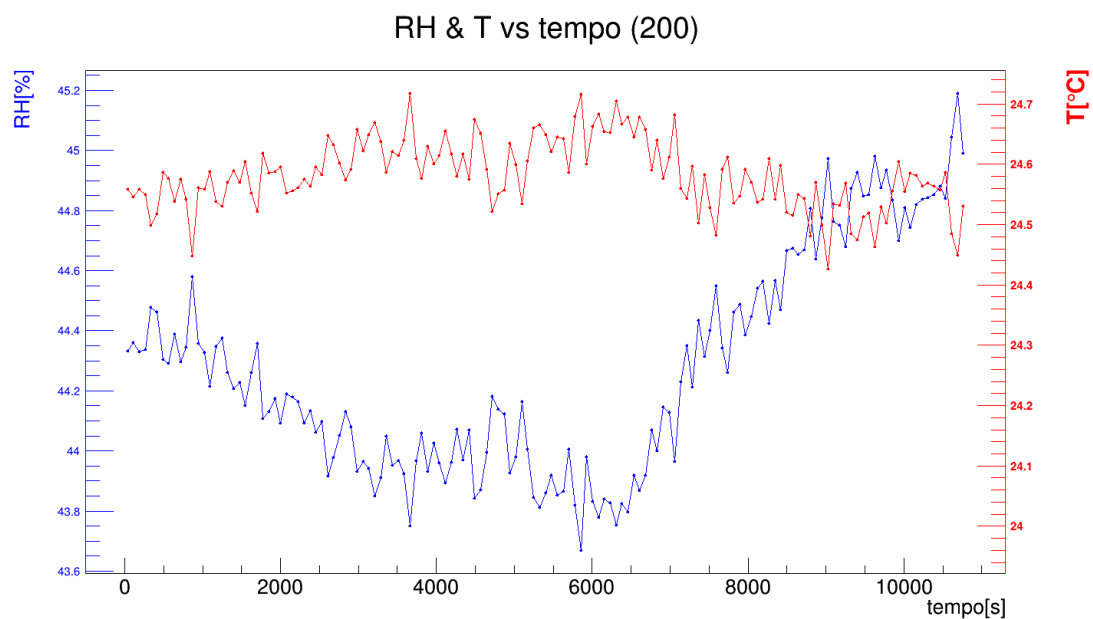


Figura 18: Andamento nel tempo dei valori medi su 200 punti della **temperatura** e dell'**umidità relativa** a confronto.

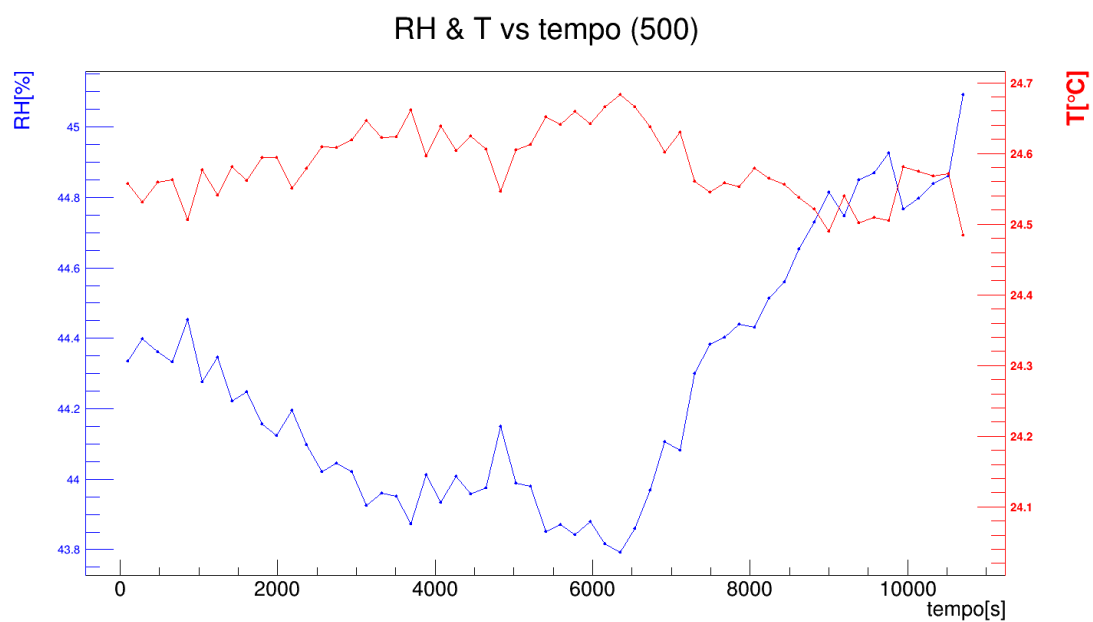


Figura 19: Andamento nel tempo dei valori medi su 500 punti della **temperatura** e dell'**umidità relativa** a confronto.

4.2 Suddivisione in sottocampioni

Il passo ulteriore per studiare localmente le grandezze è quello di suddividere l'intero *dataset* in sottocampioni e stimare le grandezze a , b e ρ per ciascun sottocampione, in modo da studiarne la compatibilità e valutarne eventualmente la media.

Le formule usate per le regressioni lineari sono le seguenti:

$$a = \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_i^2 - \sum_{i=1}^N x_i \sum_{i=1}^N x_i y_i}{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2} \quad (19)$$

$$b = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2} \quad (20)$$

$$\rho = \frac{\sum_{i=1}^N (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (21)$$

Definendo:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (y_i - a - b x_i)^2}{N - 2}} \quad (22)$$

Si ottengono le varianze:

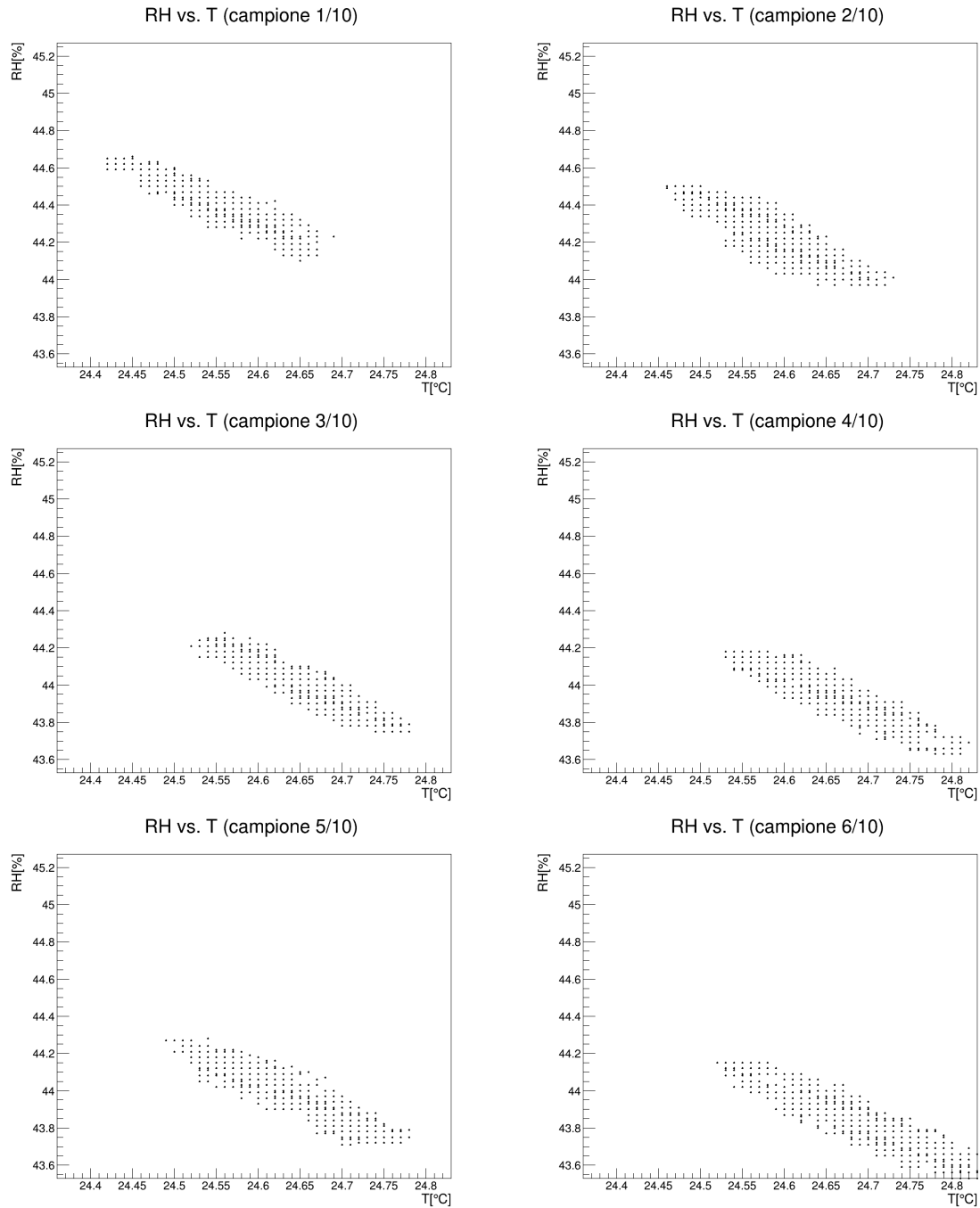
$$\text{Var}(a) = \frac{\sigma^2}{N} \frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N x_i^2 - N \bar{x}^2} \quad (23)$$

$$\text{Var}(b) = \frac{\sigma^2}{\sum_{i=1}^N x_i^2 - N \bar{x}^2} \quad (24)$$

$$\text{Var}(\rho) = \frac{1 - \rho^2}{\sqrt{N - 3}} \quad (25)$$

4.2.1 10 sottocampioni

Sono riportati di seguito gli scatter plot di ciascun sottocampione seguiti dalle stime dei parametri ottenuti tramite regressione lineare. La consistenza di tali valori verrà poi analizzata mediante un test d'ipotesi.



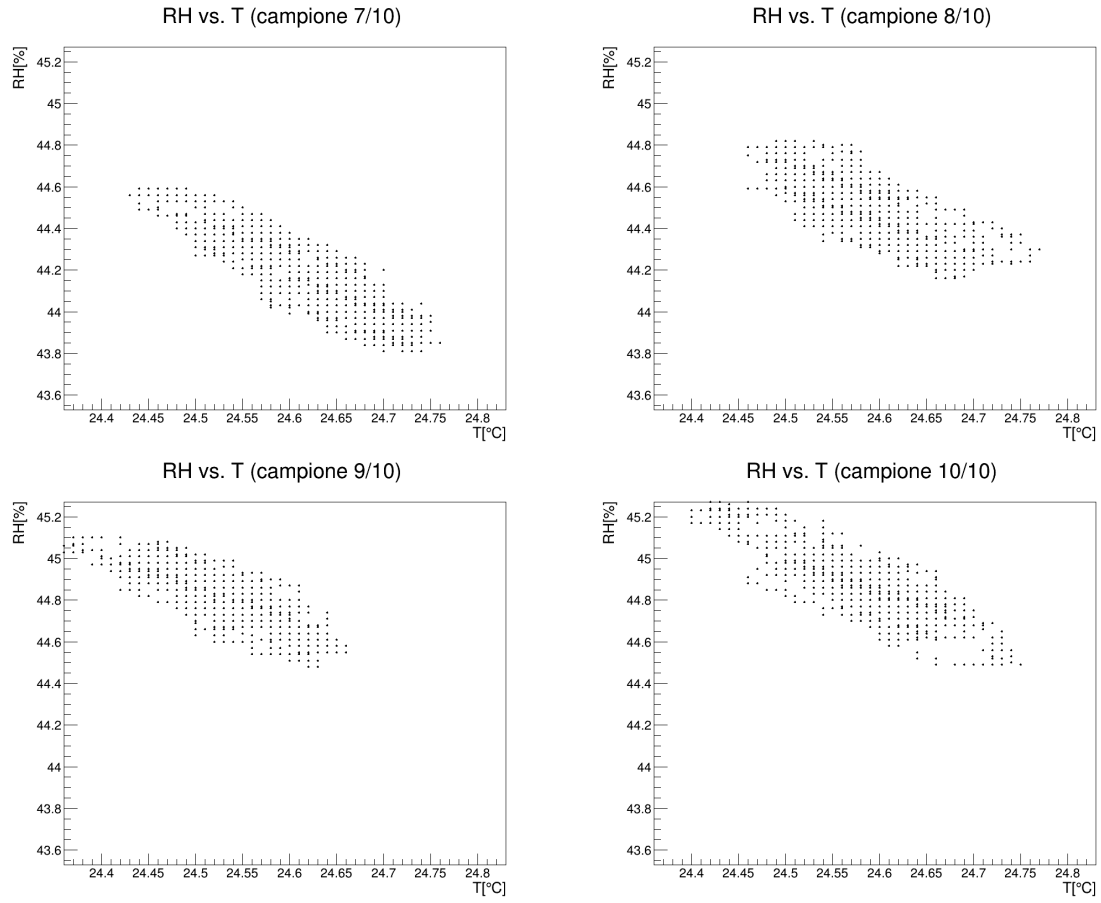


Figura 20: Scatter plot di RH e T per ciascun sottocampione.

#	$a(\%)$	$\sigma_a(\%)$	$b(\%/^{\circ}C)$	$\sigma_b(\%/^{\circ}C)$	ρ	σ_{ρ}
1	91.2	0.4	-1.91	0.02	-0.916	0.007
2	94.4	0.6	-2.04	0.02	-0.85	0.01
3	96.0	0.4	-2.11	0.02	-0.932	0.007
4	89.2	0.4	-1.83	0.02	-0.912	0.008
5	95.8	0.3	-2.10	0.01	-0.949	0.006
6	89.3	0.4	-1.84	0.02	-0.908	0.008
7	107.4	0.5	-2.57	0.02	-0.915	0.008
8	93.9	0.7	-2.01	0.03	-0.80	0.01
9	91.2	0.6	-1.89	0.02	-0.82	0.01
10	95.4	0.5	-2.06	0.02	-0.865	0.009

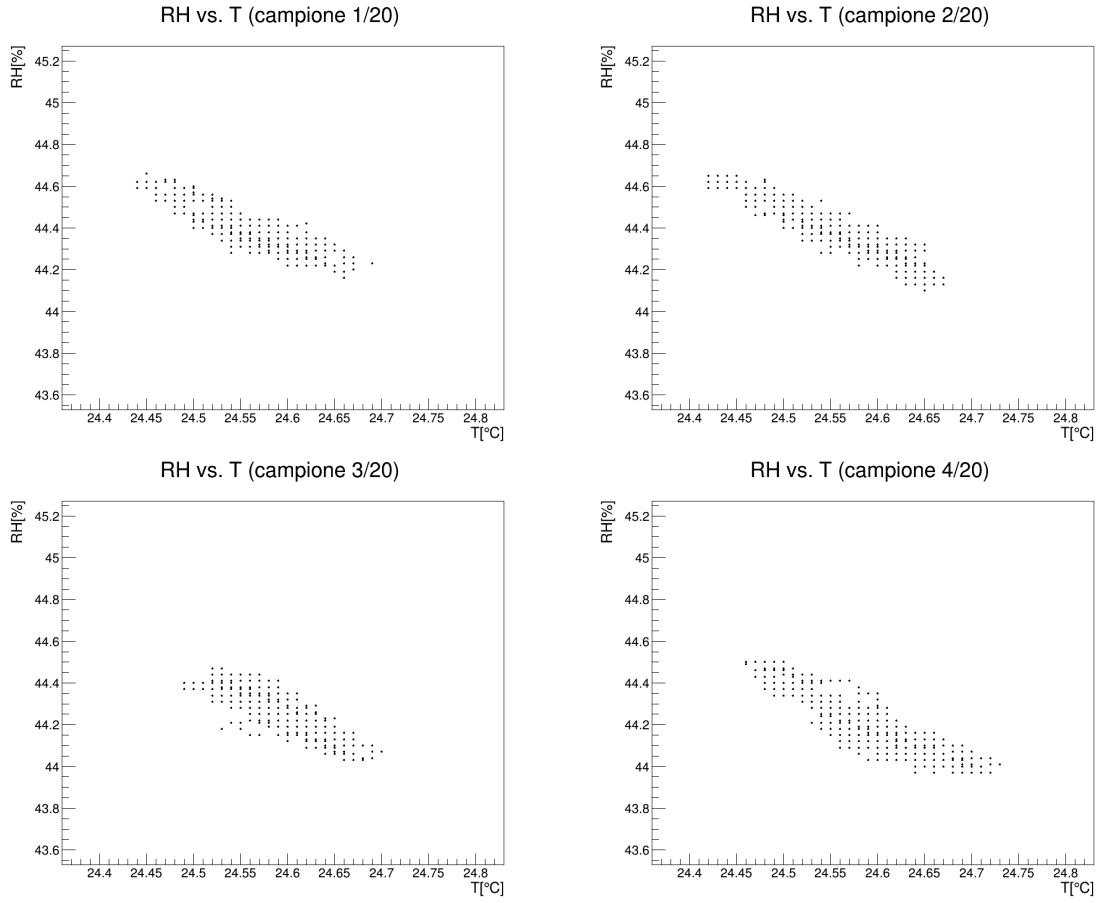
Tabella 1: Parametri stimati dai fit per i 10 sottocampioni

Le migliori stime per a , b e ρ si sono ottenute mediando i 10 valori ottenuti:

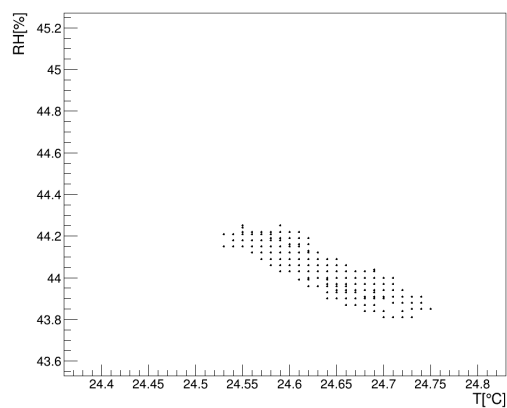
$$\begin{aligned}\bar{a} &= (94 \pm 2) \% \\ \bar{b} &= (-2.07 \pm 0.07) \%/^{\circ}C \\ \bar{\rho} &= (-0.89 \pm 0.02)\end{aligned}$$

4.2.2 20 sottocampioni

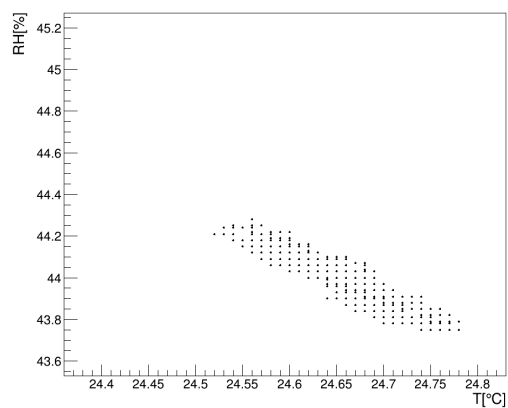
Si è ripetuto lo stesso procedimento suddividendo il *dataset* in 20 sottocampioni.



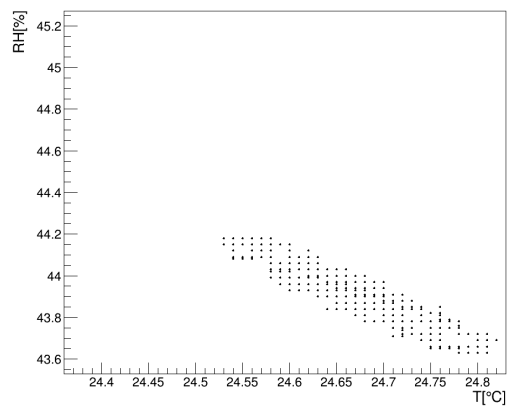
RH vs. T (campione 5/20)



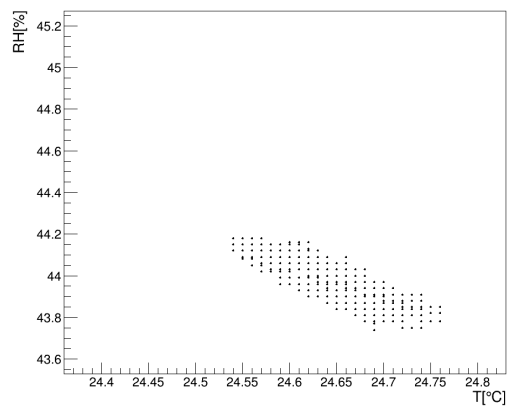
RH vs. T (campione 6/20)



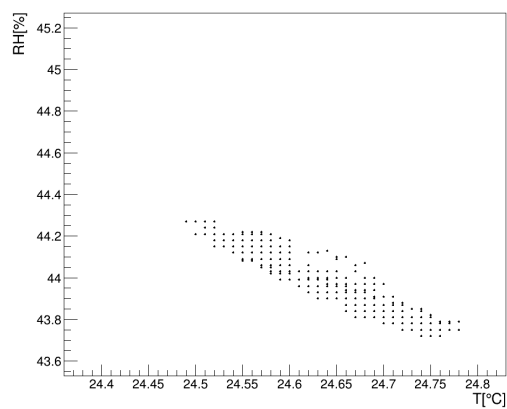
RH vs. T (campione 7/20)



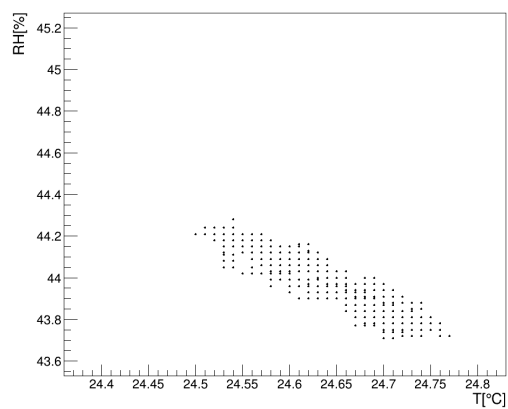
RH vs. T (campione 8/20)



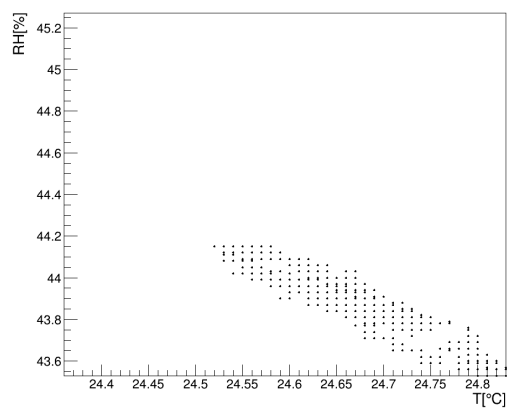
RH vs. T (campione 9/20)



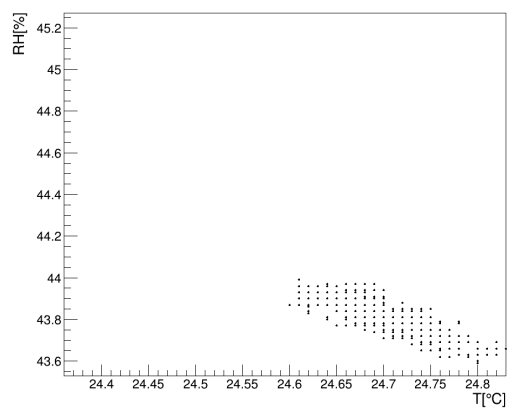
RH vs. T (campione 10/20)



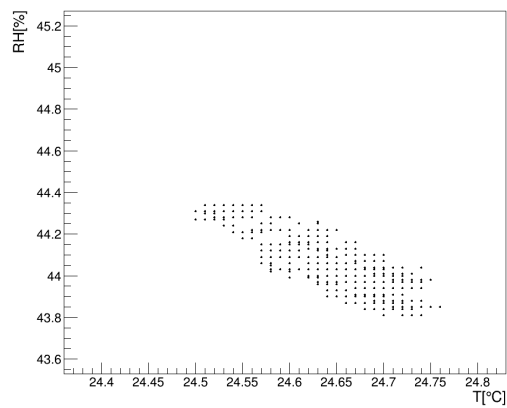
RH vs. T (campione 11/20)



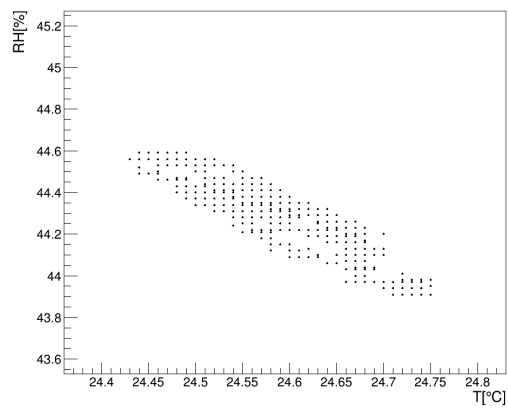
RH vs. T (campione 12/20)



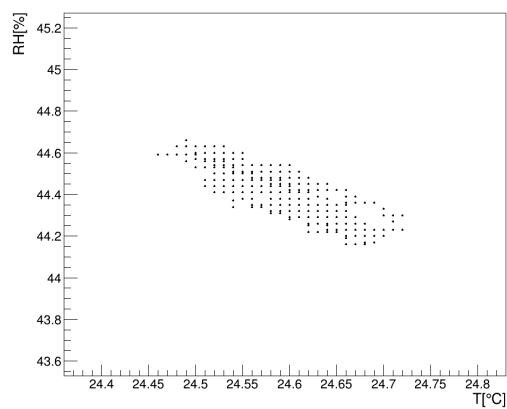
RH vs. T (campione 13/20)



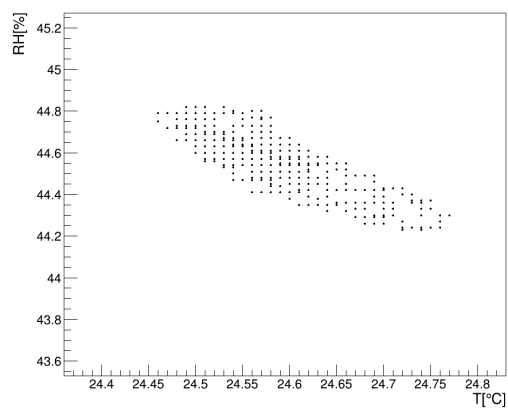
RH vs. T (campione 14/20)



RH vs. T (campione 15/20)



RH vs. T (campione 16/20)



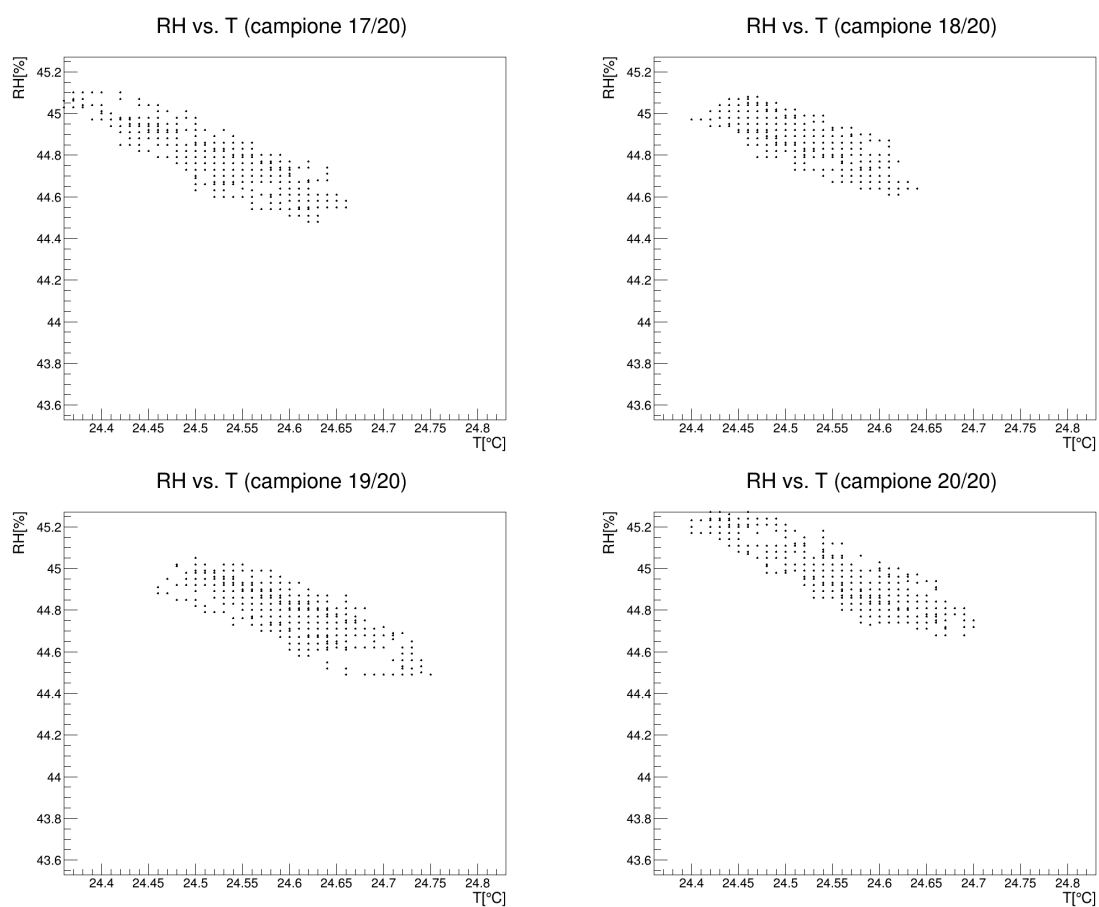


Figura 21: Scatter plot di RH e T per ciascun sottocampione

#	$a(\%)$	$\sigma_a(\%)$	$b(\%/^{\circ}C)$	$\sigma_b(\%/^{\circ}C)$	ρ	σ_{ρ}
1	89.7	0.6	-1.84	0.03	-0.888	0.012
2	92.1	0.5	-1.94	0.02	-0.934	0.009
3	92.8	0.8	-1.97	0.03	-0.854	0.014
4	93.5	0.7	-2.00	0.03	-0.891	0.012
5	92.8	0.6	-1.98	0.03	-0.902	0.011
6	97.6	0.5	-2.18	0.02	-0.937	0.009
7	89.5	0.5	-1.85	0.02	-0.933	0.009
8	88.1	0.6	-1.79	0.03	-0.883	0.012
9	94.1	0.4	-2.03	0.02	-0.951	0.008
10	96.5	0.5	-2.13	0.02	-0.940	0.009
11	92.9	0.5	-1.99	0.02	-0.931	0.010
12	80.8	0.7	-1.50	0.03	-0.829	0.015
13	92.2	0.8	-1.95	0.03	-0.857	0.014
14	96.7	0.6	-2.13	0.02	-0.919	0.010
15	91.8	0.7	-1.93	0.03	-0.872	0.013
16	88.7	0.7	-1.80	0.03	-0.865	0.013
17	89.4	0.6	-1.82	0.02	-0.893	0.012
18	89.5	0.8	-1.82	0.03	-0.824	0.015
19	85.4	0.7	-1.65	0.03	-0.821	0.015
20	93.4	0.6	-1.97	0.03	-0.899	0.012

Tabella 2: Parametri stimati dai fit per i 20 sottocampioni

Analogamente si scrivono le stime per i parametri e la correlazione:

$$\begin{aligned}\bar{a} &= (91.4 \pm 0.9) \% \\ \bar{b} &= (-1.91 \pm 0.04) \%/^{\circ}C \\ \bar{\rho} &= (-0.89 \pm 0.01)\end{aligned}$$

4.2.3 Test d'ipotesi

Per poter valutare la consistenza dei risultati ottenuti è opportuno effettuare dei test d'ipotesi. Si prende innanzitutto in considerazione il coefficiente di correlazione ρ ottenuto dai 10 sottocampioni. Dalla Figura 22 si evince che i valori di ρ sono ragionevolmente concentrati attorno alla media $\bar{\rho}$.

Vale la pena, quindi, effettuare un test che verifichi l'ipotesi nulla H_0 di non correlazione. A tale scopo si introduce la variabile casuale:

$$t_{\rho} = \frac{\rho\sqrt{N-2}}{\sqrt{1-\rho^2}} \quad (26)$$

con $N = 2878$. Questa variabile è distribuita secondo una t di Student ad $N - 2$ gradi di libertà:

$$\mathbb{E}(t_{\rho}) = 0 \quad \text{e} \quad \text{Var}(t_{\rho}) \simeq 1 \quad (27)$$

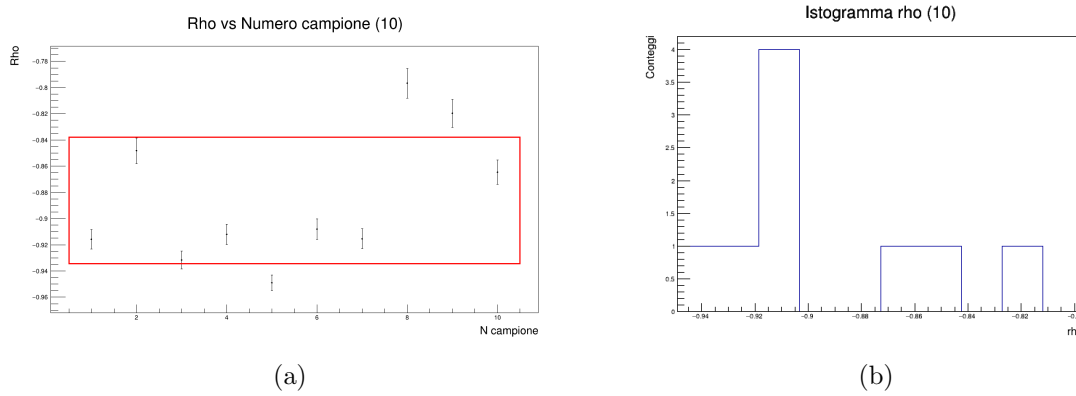


Figura 22: Scatter plot (a) ed istogramma (b) di ρ per 10 sottocampioni.

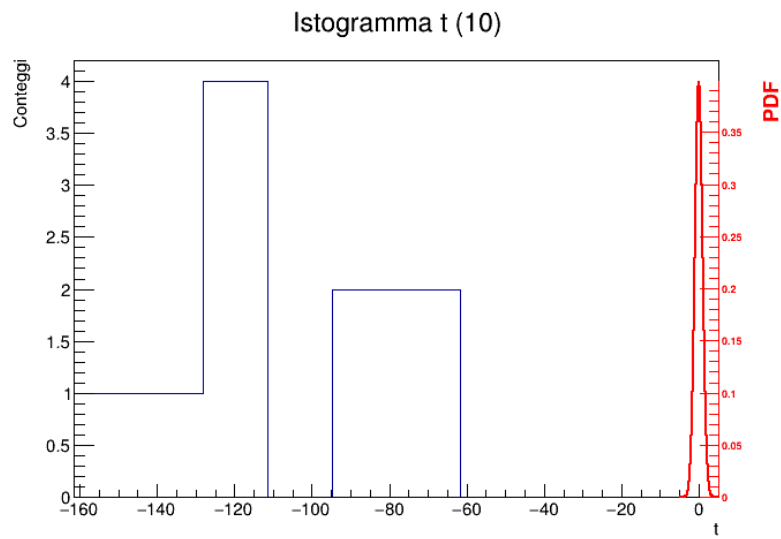


Figura 23: Istogramma di t_ρ per 10 sottocampioni a confronto con la sua distribuzione attesa.

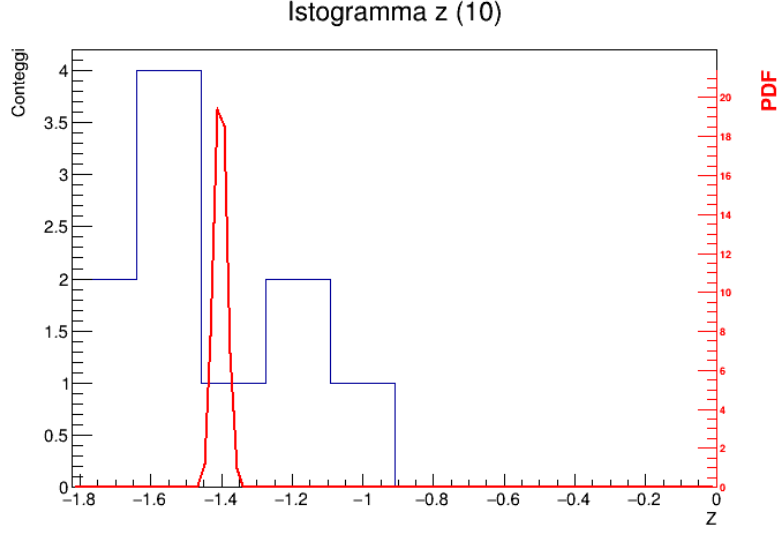


Figura 24: Istogramma di $Z(\rho)$ per 10 sottocampioni a confronto con la sua distribuzione attesa.

Scegliendo una significatività $\alpha = 0.05$, si osserva in Figura 23 che nessuno degli intervalli ricade nell'intervallo di confidenza sotteso dalla distribuzione. Va rigettata, come ci si aspettava, l'ipotesi di non correlazione tra le grandezze misurate.

Si verifica ora la consistenza dei valori di ρ con la loro media, utilizzando come statistica di test la distribuzione di Fisher. Si definisce quindi:

$$Z(\rho) = \frac{1}{2} (\ln(1 + \rho) - \ln(1 - \rho)) \quad (28)$$

La distribuzione di Fisher tende, per elevati gradi di libertà, ad una gaussiana $G(Z(\bar{\rho}), \sigma_Z)$ con i seguenti parametri:

$$\mathbb{E}(Z(\rho)) = Z(\bar{\rho}) \quad \text{e} \quad \text{Var}(Z) = \sigma_Z^2 = \frac{1}{N-3} \quad (29)$$

In tabella 3 sono riportati i valori della variabile di Fisher per 10 sottocampioni ed in Figura 24 vengono confrontati con la loro distribuzione attesa. Scegliendo per la significatività sempre un valore $\alpha = 0.05$, considerando che $\sigma_Z = 0.02$ e $Z(\bar{\rho}) = -1.40$, si evince che nessuno dei valori ricade all'interno dell'intervallo di confidenza.

Al fine di valutare la compatibilità delle varie stime di ρ si studia la variabile casuale:

$$\delta = \frac{Z(\rho_i) - Z(\rho_j)}{\sqrt{\frac{1}{N_i-3}} \sqrt{\frac{1}{N_j-3}}} \quad (30)$$

la quale segue una gaussiana $G(0, 1)$. In Figura 25 è riportata la distribuzione dei tale variabile per tutte le possibili coppie di ρ dei 10 sottocampioni. É evidente che la quasi totalità dei valori ricade all'infuori dell'intervallo di confidenza.

#	$Z(\rho)$
1	-1.56
2	-1.25
3	-1.67
4	-1.54
5	-1.82
6	-1.52
7	-1.56
8	-1.09
9	-1.16
10	-1.31

Tabella 3: Valori della variabile $Z(\rho)$ di Fisher per 10 sottocampioni

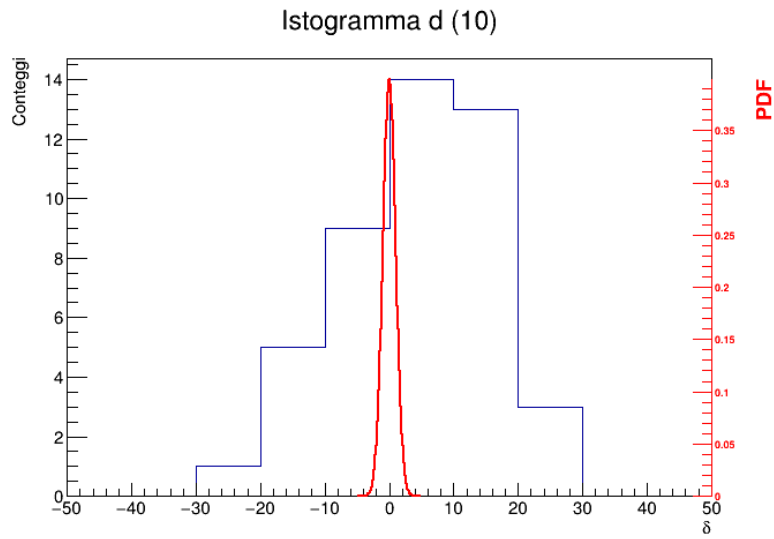


Figura 25: Istogramma della δ per 10 sottocampioni a confronto con la sua distribuzione attesa.

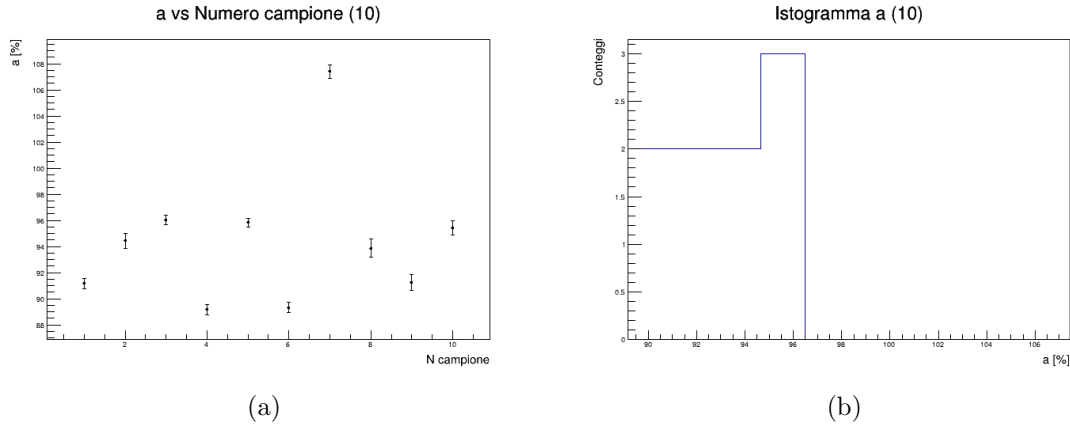


Figura 26: Scatter plot (a) ed istogramma (b) di a per 10 sottocampioni.

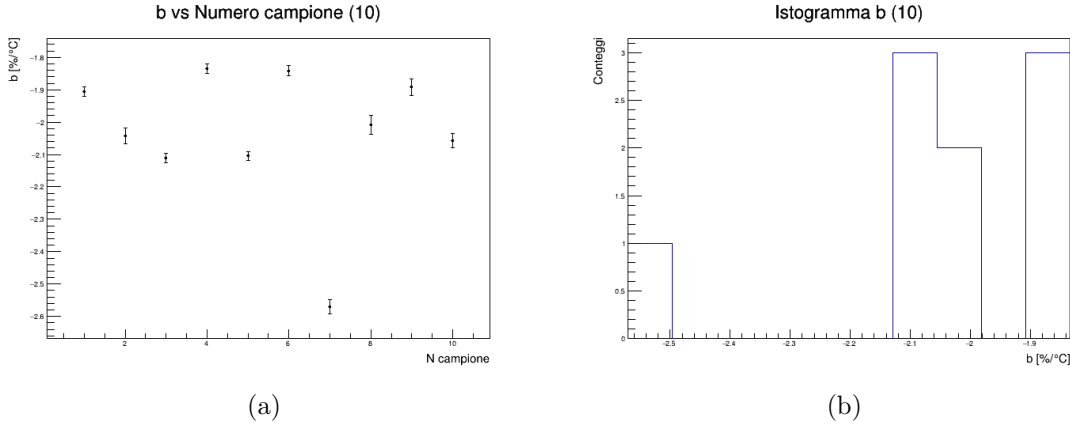


Figura 27: Scatter plot (a) ed istogramma (b) di b per 10 sottocampioni.

Per i parametri a e b viene fatto un confronto con una t di Student ad $N - 2$ gradi di libertà così da verificare l'ipotesi secondo cui la migliore stima per tali parametri si la loro media pesata. Si definisce quindi:

$$t_a = \frac{a - \bar{a}}{\sigma_a} \quad t_b = \frac{b - \bar{b}}{\sigma_b} \quad (31)$$

Nelle Figure 26 e 27 vengono riportati i valori di a e b ed in Figura 28 viene mostrato il T-test.

Il motivo delle inconsistenze evidenziate fin ora va probabilmente ricercato nella presenza di un errore sistematico che è possibile stimare tramite la variabile χ^2 :

$$\chi_x^2 = \sum_i \frac{(x_i - \bar{x})^2}{\sigma_{x_i}^2} \quad (32)$$

dove $x = a, b, \rho$. Si trovano quindi i seguenti valori:

$$\chi_a^2 = 1131 \quad \chi_b^2 = 1122 \quad \chi_\rho^2 = 331$$

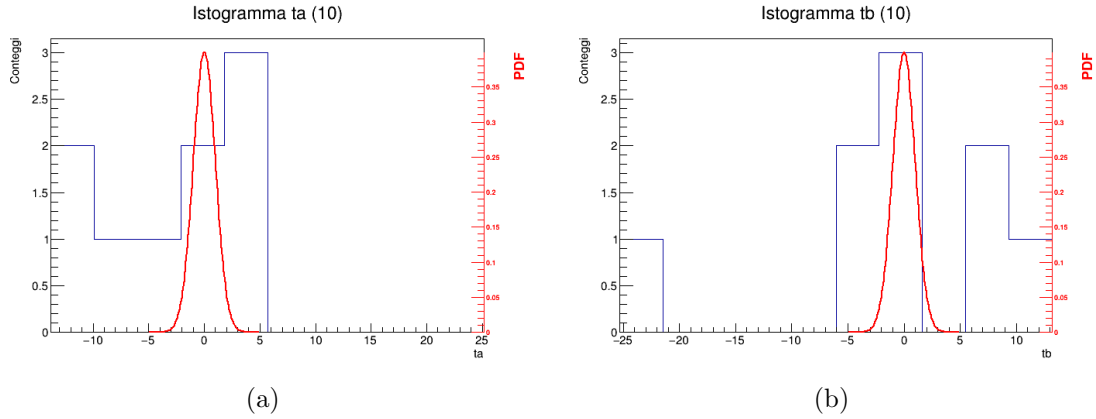


Figura 28: Distribuzioni di t_a e t_b per 10 sottocampioni con le loro pdf attese.

Il valore di confidenza è pari a $\chi^2_{\text{lim}} = 16.92$, con 9 gradi di libertà e una significatività pari a $\alpha = 0.05$, evidentemente ben distante dai valori ottenuti. Tutte le misure sono quindi affette da errore sistematico. La stima di tali errori ha portato i seguenti risultati:

$$\sigma_a^{\text{sys}} = 3\% \quad \sigma_b^{\text{sys}} = 0.12\%/^{\circ}C \quad \sigma_{\rho}^{\text{sys}} = 0.03$$

L'intera analisi viene analogamente ripetuta per la suddivisione in 20 sottocampioni ($N = 1439$). Nelle figure 29, 30 e 31 sono riportati le distribuzioni dei parametri ottenuti dai fit. Il grafico in Figura 32 mostra che anche in questo caso l'ipotesi di non correlazione va rigettata.

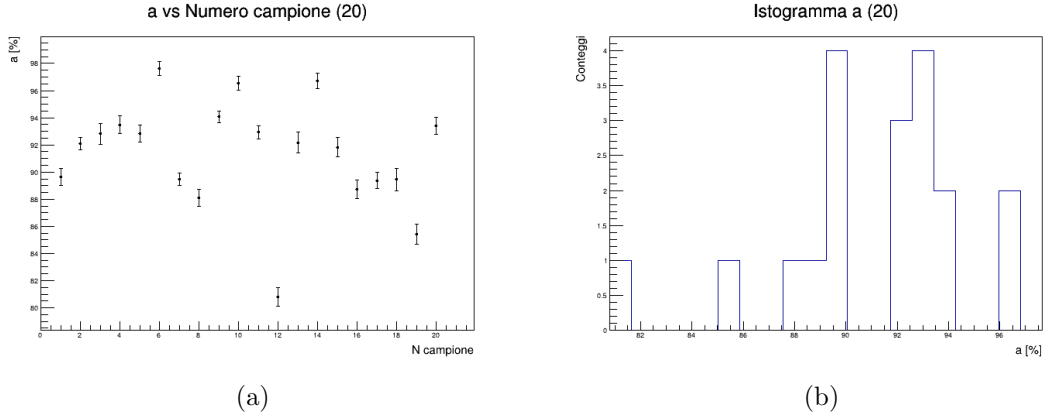


Figura 29: Scatter plot (a) ed istogramma (b) di a per 20 sottocampioni.

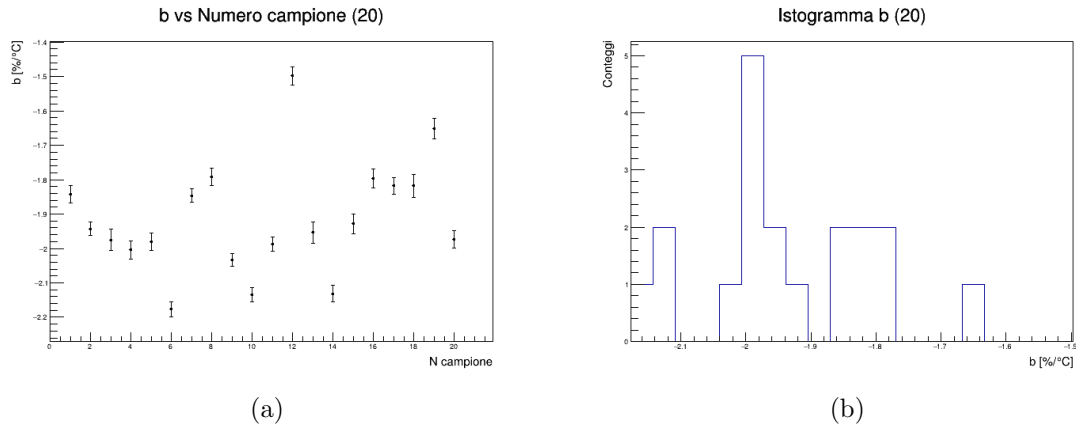


Figura 30: Scatter plot (a) ed istogramma (b) di b per 20 sottocampioni.

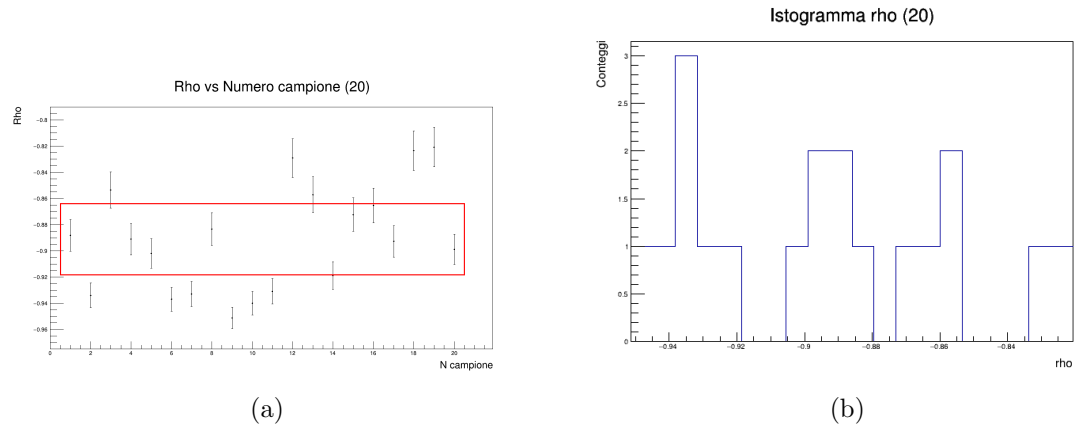


Figura 31: Scatter plot (a) ed istogramma (b) di ρ per 20 sottocampioni.

In tabella 4 sono ripetuti i calcoli per la variabile di Fisher, stavolta con $Z_{\bar{\rho}} = -1.4$ e $\sigma_Z = 0.03$.

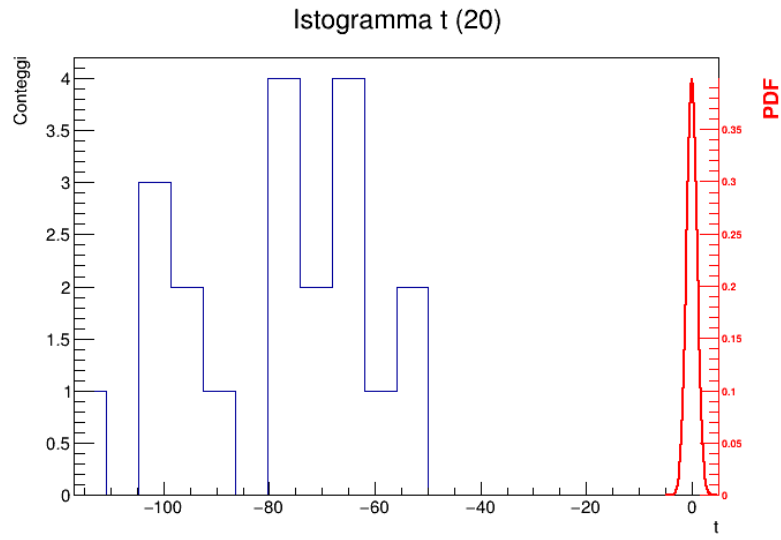


Figura 32: Istogramma di t_ρ per 20 sottocampioni a confronto con la sua distribuzione attesa.

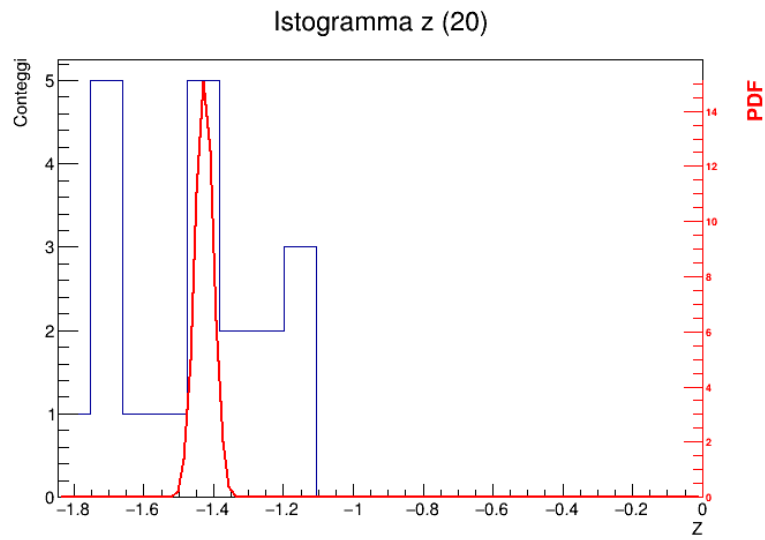


Figura 33: Istogramma di $Z(\rho)$ per 20 sottocampioni a confronto con la sua distribuzione attesa.

#	$Z(\rho)$
1	-1.41
2	-1.69
3	-1.27
4	-1.43
5	-1.48
6	-1.71
7	-1.68
8	-1.39
9	-1.84
10	-1.74
11	-1.66
12	-1.19
13	-1.28
14	-1.58
15	-1.34
16	-1.31
17	-1.44
18	-1.17
19	-1.16
20	-1.47

Tabella 4: Valori della variabile $Z(\rho)$ di Fisher per 20 sottocampioni

Sono solo 4 i valori di Z che ricadono nell'intervallo di confidenza (Fig. 33), evidenza ulteriore del fatto che la misura è affetta da errore sistematico. La compatibilità tra le coppie di ρ viene nuovamente studiata tramite la variabile casuale δ in Figura 34. Ripetendo il test con la statistica χ^2 a 19 gradi di libertà ed $\alpha = 0.05$, ossia un valore di soglia $\chi^2_{\text{lim}} = 31.14$, si ottengono i seguenti risultati:

$$\chi_a^2 = 816 \quad \chi_b^2 = 845 \quad \chi_\rho^2 = 254$$

e, di conseguenza, i seguenti errori sistematici:

$$\sigma_a^{\text{sys}} = 2\% \quad \sigma_b^{\text{sys}} = 0.09\%/^{\circ}C \quad \sigma_\rho^{\text{sys}} = 0.02$$

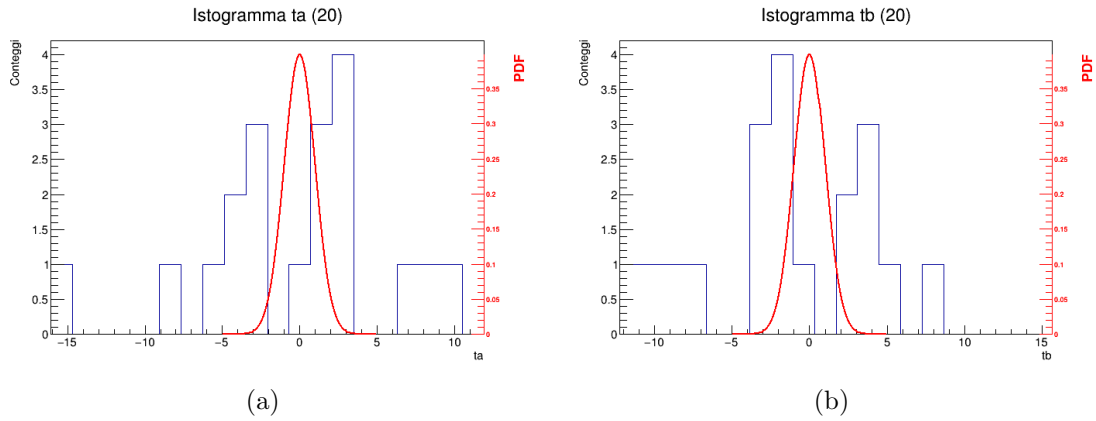


Figura 35: Distribuzioni di t_a e t_b per 20 sottocampioni con le loro pdf attese.

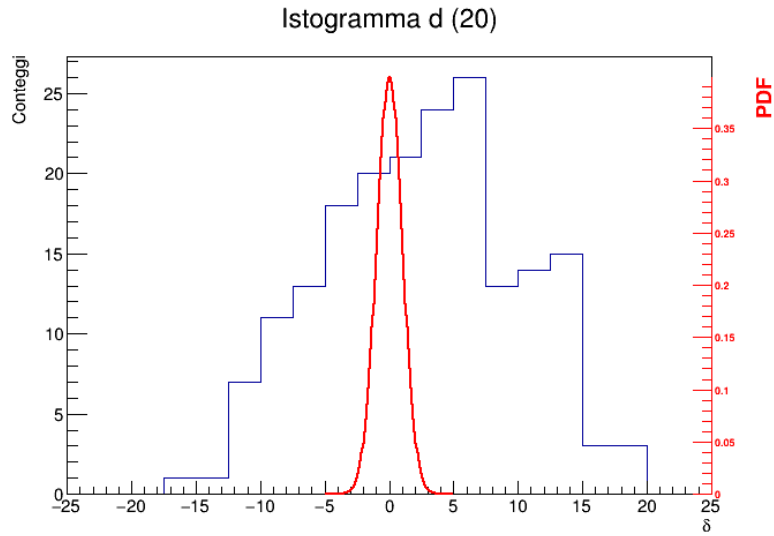


Figura 34: Istogramma della δ per 20 sottocampioni a confronto con la sua distribuzione attesa.

Per le variabili a e b è stato ripetuto il test con la t di Student come mostrato in Figura 35.

4.3 Sequenze temporali

Un'indagine ulteriore che si può fare è quella di andare a considerare campioni che contengono pochi punti sperimentali, racchiusi in intervalli di tempo molto brevi. Così facendo si ignora l'evoluzione temporale del sistema ed è possibile apprezzare la randomicità della misura. Nello specifico sono stati considerati i primi 1020 punti, ossia i primi 7 minuti della presa dati, i quali sono stati a loro volta suddivisi in campioni da 5/10/15/20 punti.

Come anticipato, le distribuzioni dei parametri ottenuti dai fit su ciascuno di questi campioni evidenziano un andamento casuale (Fig. da 36 a 47). Inoltre, i valori di b e di ρ sono distribuiti attorno a 0.

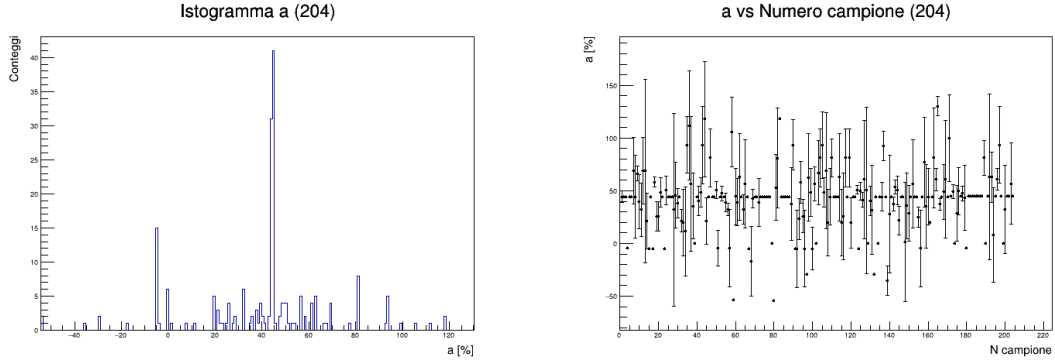


Figura 36: Scatter plot ed istogramma di a per sequenze di 5 punti.

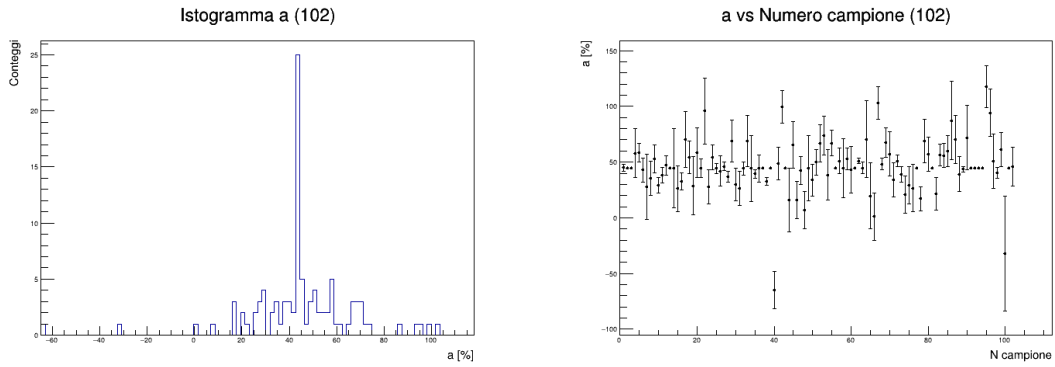


Figura 37: Scatter plot ed istogramma di a per sequenze di 10 punti.

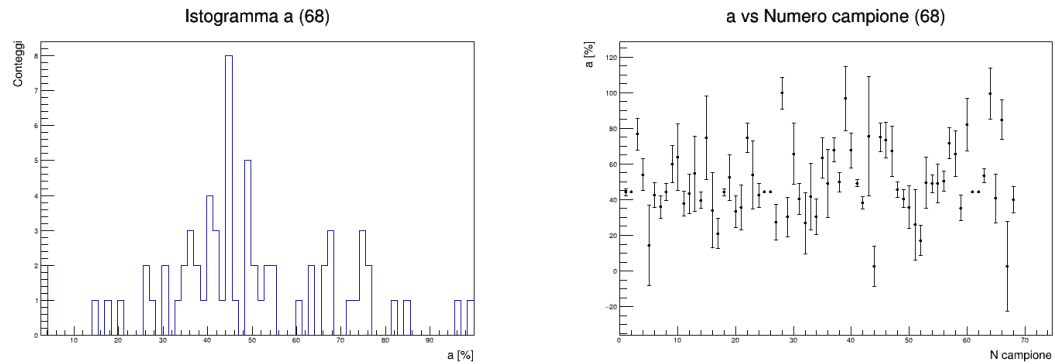


Figura 38: Scatter plot ed istogramma di a per sequenze di 15 punti.

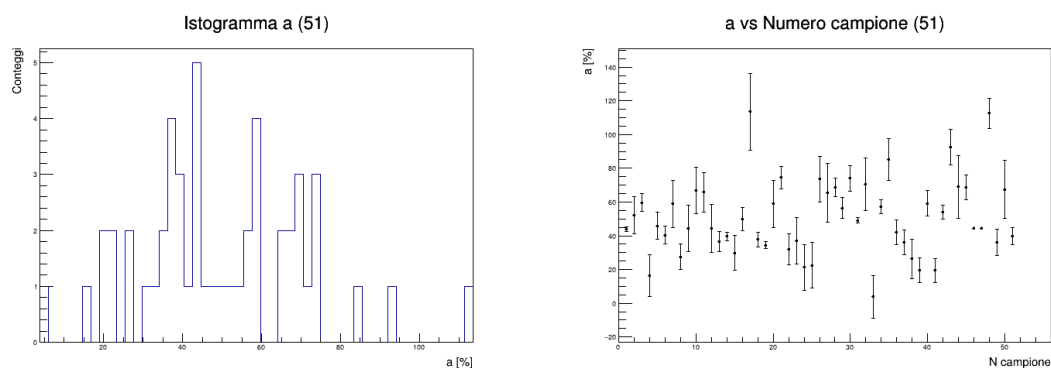


Figura 39: Scatter plot ed istogramma di a per sequenze di 20 punti.

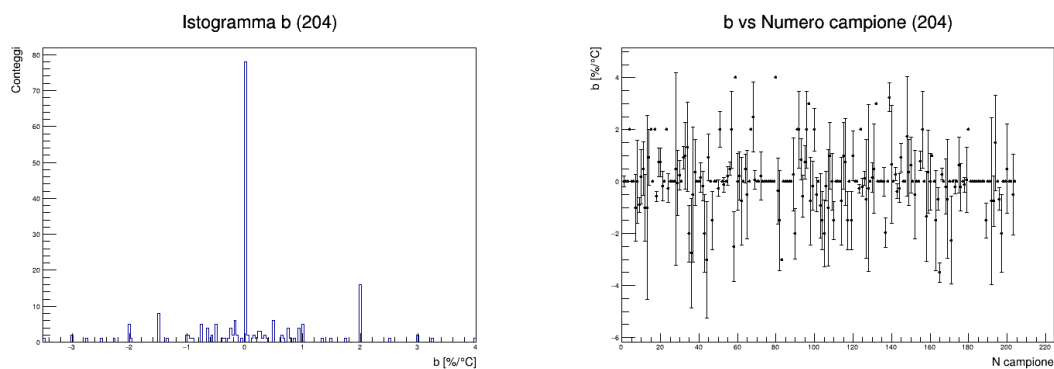


Figura 40: Scatter plot ed istogramma di b per sequenze di 5 punti.

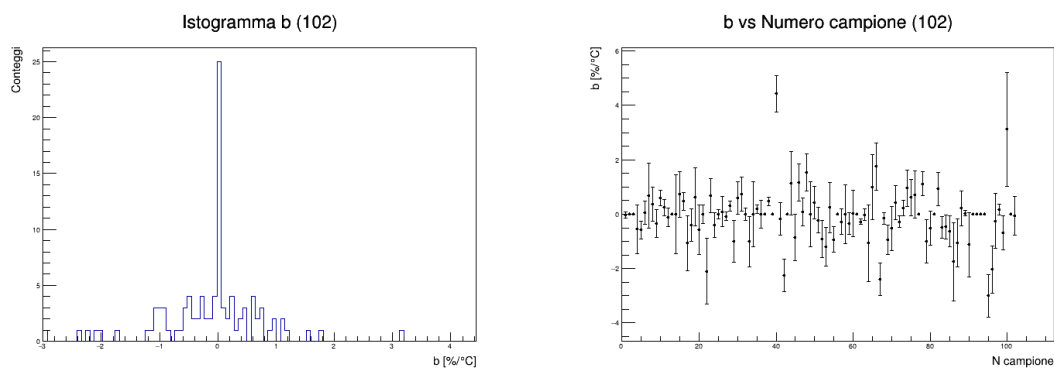


Figura 41: Scatter plot ed istogramma di b per sequenze di 10 punti.

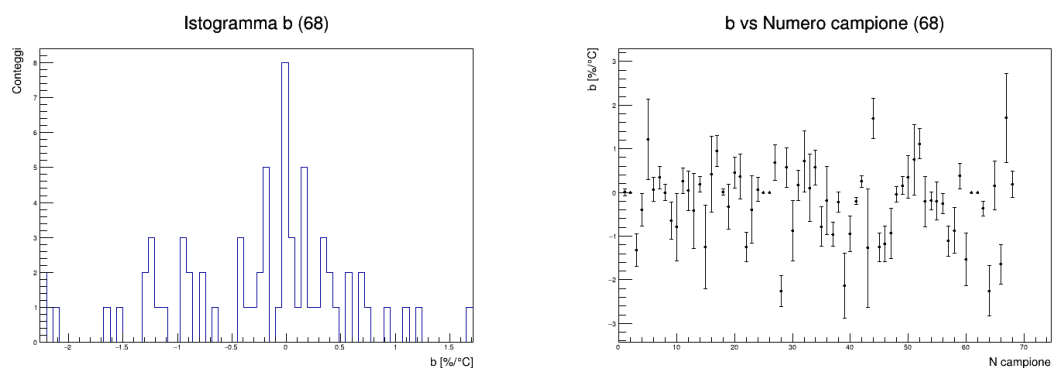


Figura 42: Scatter plot ed istogramma di b per sequenze di 15 punti.

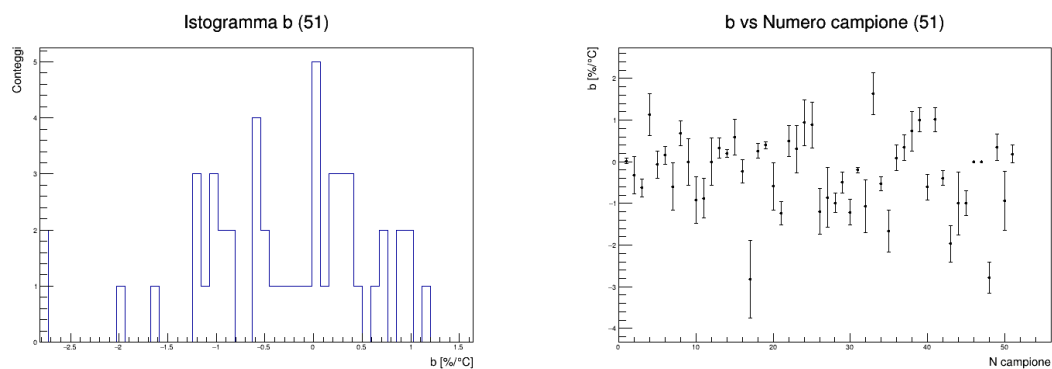


Figura 43: Scatter plot ed istogramma di b per sequenze di 20 punti.

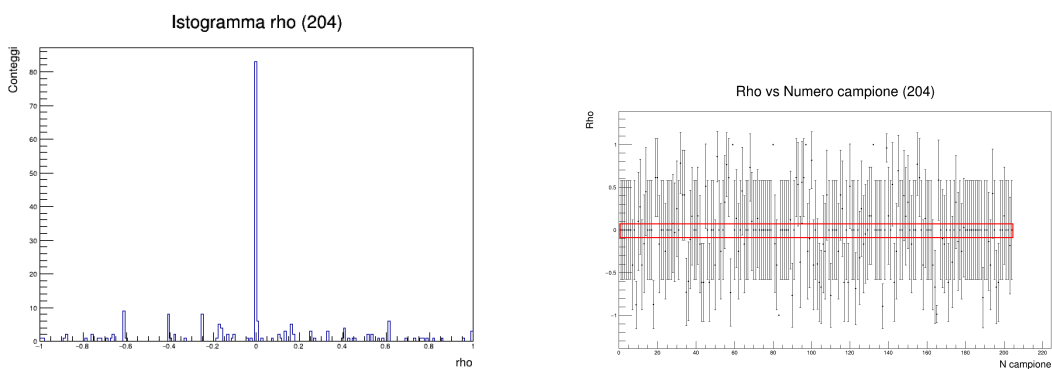


Figura 44: Scatter plot ed istogramma di ρ per sequenze di 5 punti.

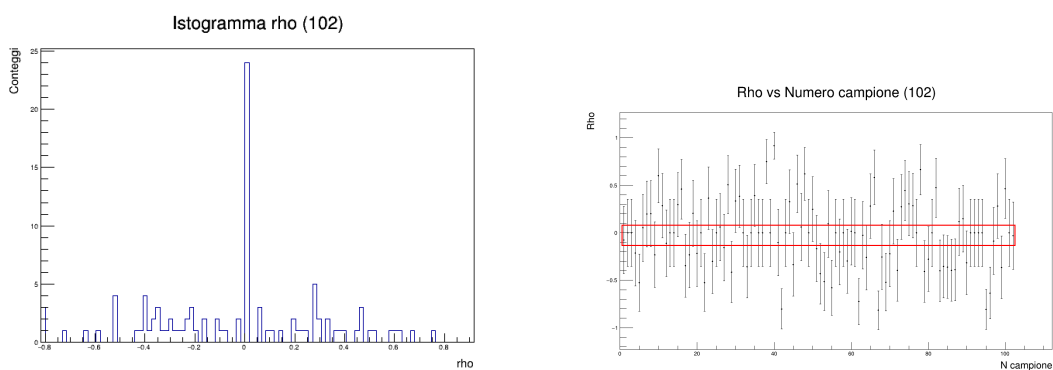


Figura 45: Scatter plot ed istogramma di ρ per sequenze di 10 punti.

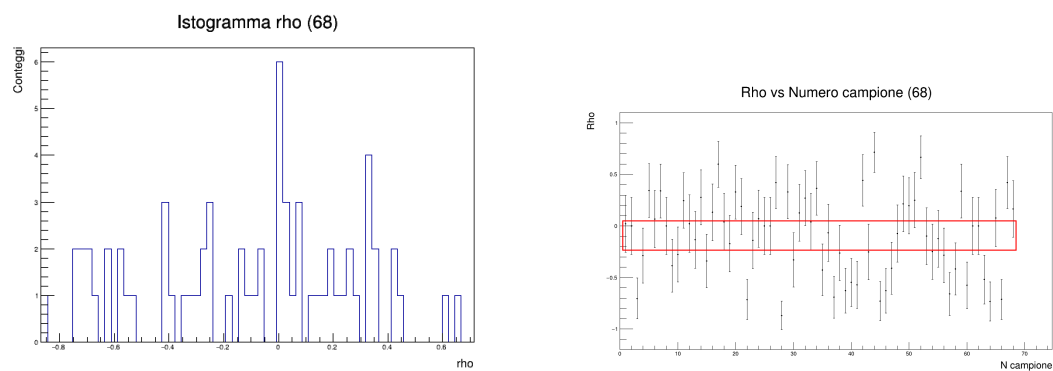


Figura 46: Scatter plot ed istogramma di ρ per sequenze di 15 punti.

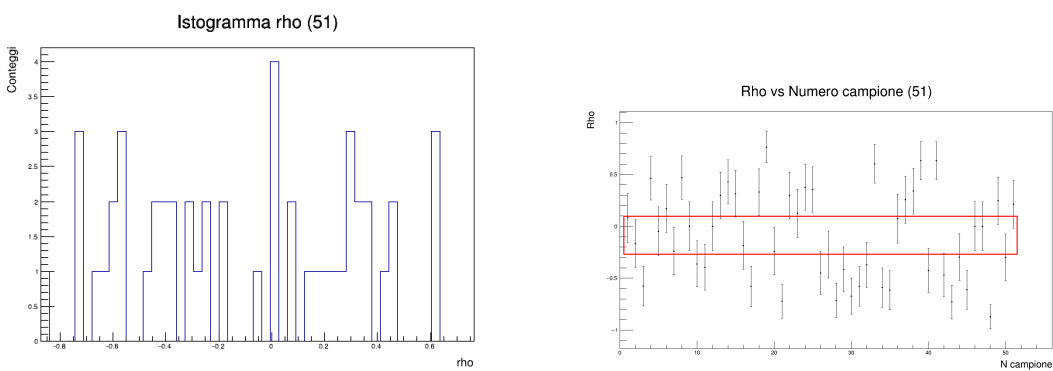


Figura 47: Scatter plot ed istogramma di ρ per sequenze di 20 punti.

Si è ripetuto il confronto con la distribuzione t di Student e della distribuzione di Fisher per la correlazione ρ . Si nota come, al diminuire del numero di punti per sottocampione, si ottiene un numero gradualmente crescente di valori compresi nei rispettivi intervalli di confidenza. Questo è un'ulteriore comprova di quanto affermato prima: considerando sequenze piccole al punto da rendere impercettibile l'evoluzione del sistema, i valori dei parametri seguono un andamento aleatorio. Va notata, tuttavia, la presenza di alcune stime di ρ particolarmente vicine al valore 1, con errori molto piccoli. Questo comportamento si ritrova anche nei valori della $Z(\rho)$ ed è particolarmente visibile nei campioni meno popolati. Ciò è dovuto al fatto che, essendo le misure di ciascun campione temporalmente molto ravvicinate, vi possono essere campioni i cui punti sono perfettamente allinati o addirittura sovrapposti (punti stazionari).

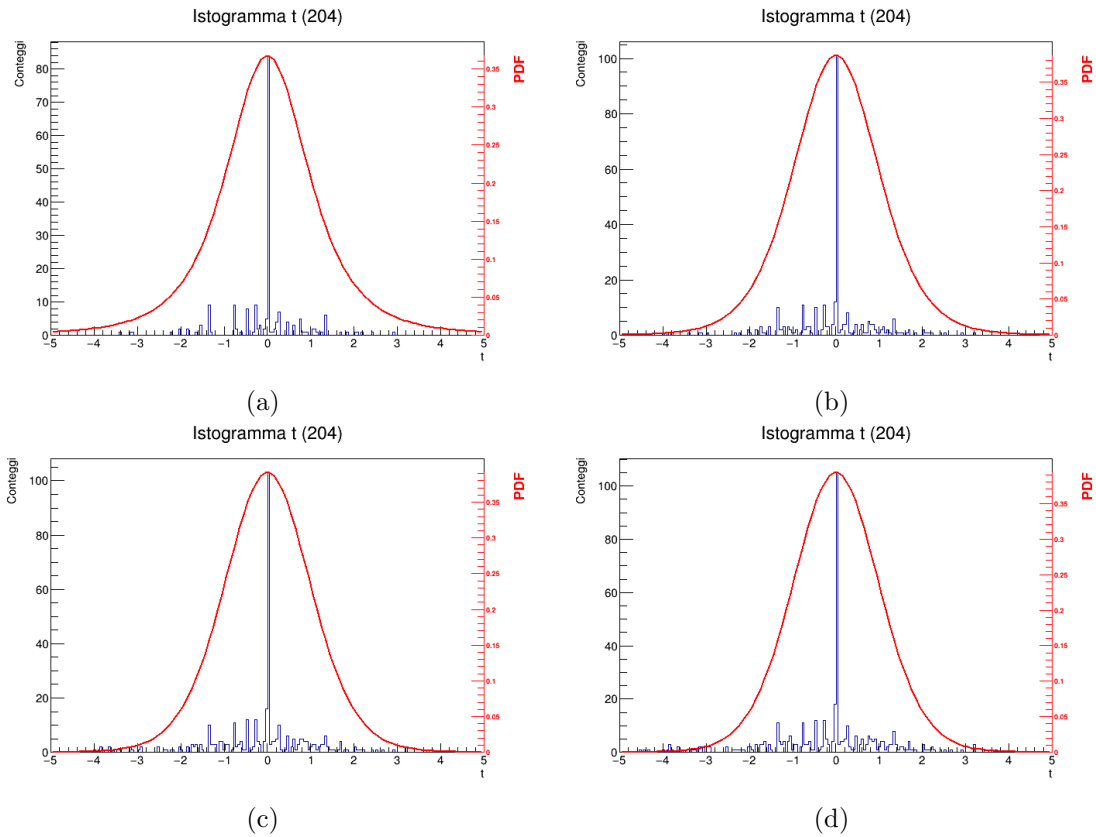


Figura 48: Istogramma di t_ρ e rispettiva pdf per 5 (a), 10 (b), 15 (c) e 20 (d) punti.

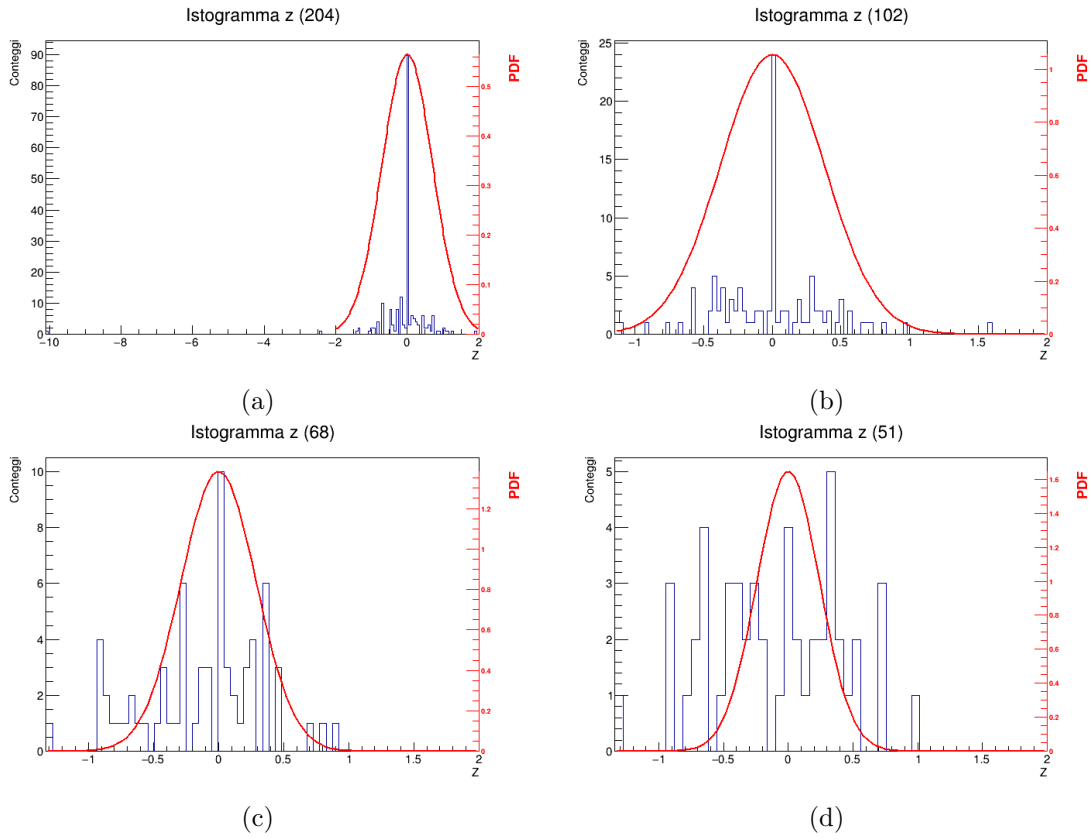


Figura 49: Istogramma di $Z(\rho)$ e rispettiva pdf per 5 (a), 10 (b), 15 (c) e 20 (d) punti.

Per i parametri a e b è stato ripetuto il test d'ipotesi con la statistica di Student come riportato nelle Figure 50 e 51.

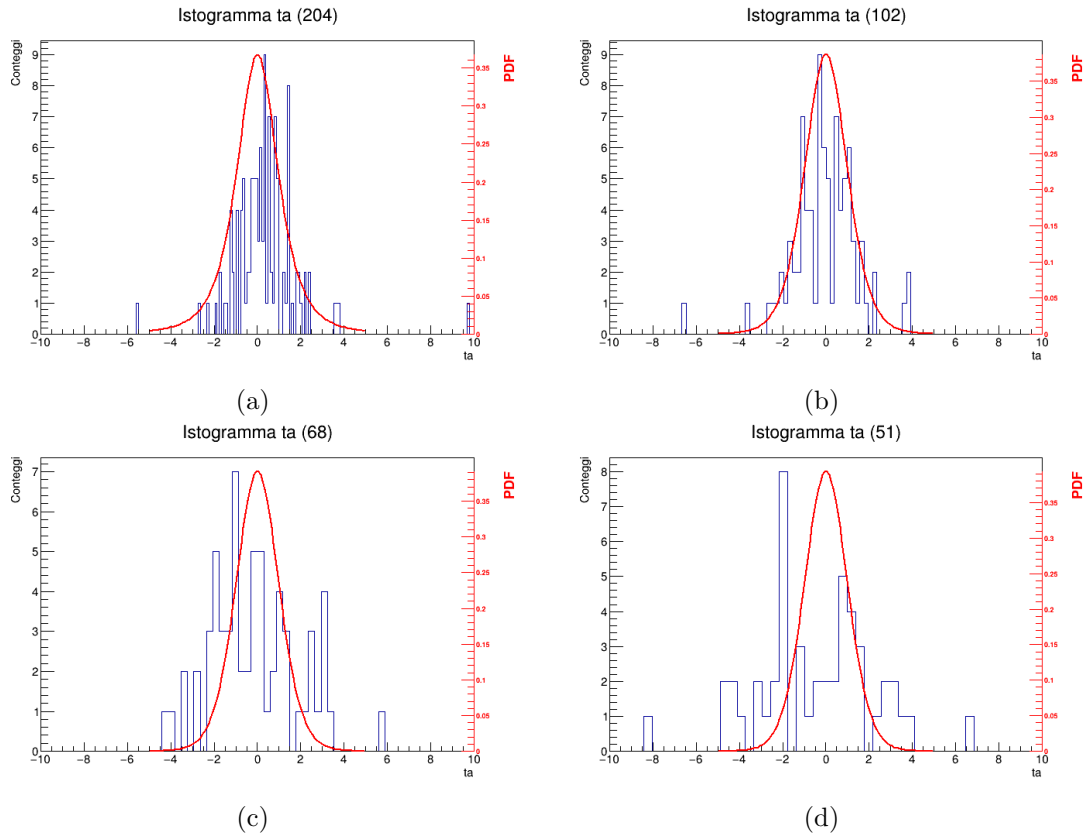


Figura 50: Istogramma di t_a e rispettiva pdf per 5 (a), 10 (b), 15 (c) e 20 (d) punti.

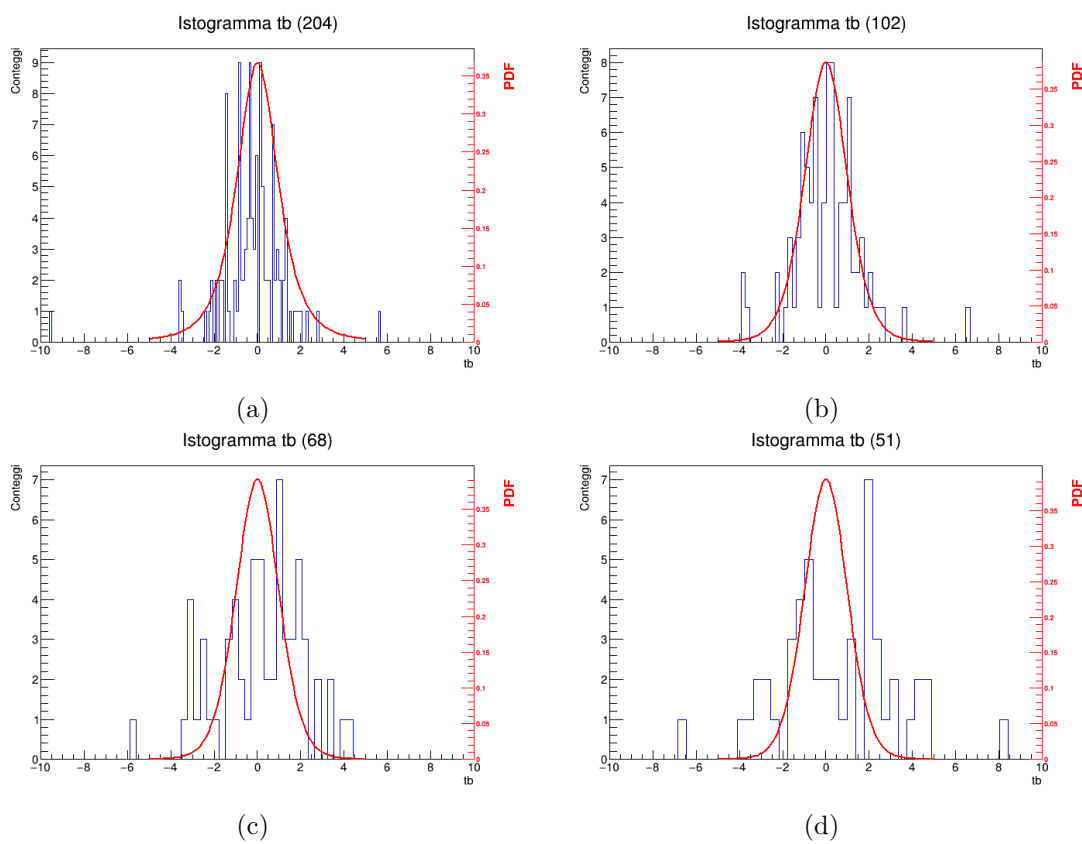


Figura 51: Istogramma di t_b e rispettiva pdf per 5 (a), 10 (b), 15 (c) e 20 (d) punti.

Conclusioni

In definitiva l'analisi del *dataset* ci consente di confermare la correlazione lineare tra umidità relativa e temperatura, nonché della loro relazione funzionale. Si sono valutate come migliori stime dei coefficienti le medie dei paramtri ottenuti dai fit sui 10 o 20 sottocampioni del set intero. Tali valori sono riportati di seguito:

$$\begin{aligned}a &= (94 \pm 2) \% \\b &= (-2.07 \pm 0.07) \% / ^\circ C \\\rho &= (-0.89 \pm 0.02)\end{aligned}$$

(per 10 sottocampioni)

$$\begin{aligned}a &= (91.4 \pm 0.9) \% \\b &= (-1.91 \pm 0.04) \% / ^\circ C \\\rho &= (-0.89 \pm 0.01)\end{aligned}$$

(per 20 sottocampioni)

Le stime sono tra loro compatibili, ma non lo sono con il valore teorico atteso.

Appendice A: Tabella dei primi 100 eventi

#	Data	Tempo dalla mezzanotte (s)	Tempo dall' inizio (s)	RH_{RAW}	$RH(\%)$	T_{RAW}	$T(\%)$
1	2020 12 3	50408.00	0.801669	1346	44.41	6415	24.55
2	2020 12 3	50408.00	1.202116	1346	44.41	6414	24.54
3	2020 12 3	50408.00	1.602578	1346	44.41	6415	24.55
4	2020 12 3	50409.00	2.003051	1346	44.41	6412	24.52
5	2020 12 3	50409.00	2.403529	1346	44.40	6414	24.54
6	2020 12 3	50410.00	2.804016	1346	44.41	6414	24.54
7	2020 12 3	50410.00	3.204479	1346	44.41	6414	24.54
8	2020 12 3	50410.00	3.604698	1346	44.41	6413	24.53
9	2020 12 3	50411.00	4.005212	1346	44.41	6413	24.53
10	2020 12 3	50411.00	4.405678	1346	44.41	6414	24.54
11	2020 12 3	50412.00	4.806189	1346	44.41	6415	24.55
12	2020 12 3	50412.00	5.206589	1346	44.41	6414	24.54
13	2020 12 3	50412.00	5.607147	1346	44.41	6414	24.54
14	2020 12 3	50413.00	6.007548	1346	44.41	6415	24.55
15	2020 12 3	50413.00	6.007548	1346	44.41	6415	24.55
16	2020 12 3	50413.00	6.407899	1346	44.41	6415	24.55
17	2020 12 3	50414.00	6.808465	1346	44.41	6415	24.55
18	2020 12 3	50414.00	7.208844	1346	44.41	6415	24.55
19	2020 12 3	50414.00	7.609052	1346	44.41	6415	24.55
20	2020 12 3	50415.00	8.009619	1346	44.41	6415	24.55
21	2020 12 3	50415.00	8.410081	1346	44.41	6416	24.56
22	2020 12 3	50416.00	8.810502	1346	44.41	6415	24.55
23	2020 12 3	50416.00	9.211041	1346	44.41	6415	24.55
24	2020 12 3	50416.00	9.611555	1346	44.41	6414	24.54
25	2020 12 3	50417.00	10.012165	1346	44.41	6414	24.54
26	2020 12 3	50417.00	10.412381	1346	44.41	6415	24.55
27	2020 12 3	50418.00	10.812929	1346	44.41	6414	24.54
28	2020 12 3	50418.00	11.213407	1346	44.41	6415	24.55
29	2020 12 3	50418.00	11.613958	1346	44.41	6414	24.54
30	2020 12 3	50419.00	12.014476	1346	44.41	6415	24.55
31	2020 12 3	50419.00	12.014476	1346	44.41	6414	24.54
32	2020 12 3	50419.00	12.414765	1346	44.41	6415	24.55
33	2020 12 3	50420.00	12.814993	1346	44.41	6414	24.54
34	2020 12 3	50420.00	13.215412	1346	44.41	6415	24.55
35	2020 12 3	50420.00	13.616001	1345	44.38	6415	24.55
36	2020 12 3	50421.00	14.016245	1346	44.41	6414	24.54
37	2020 12 3	50421.00	14.416577	1345	44.37	6415	24.55
38	2020 12 3	50422.00	14.816952	1346	44.41	6416	24.56
39	2020 12 3	50422.00	15.217407	1346	44.41	6415	24.55
40	2020 12 3	50422.00	15.617880	1345	44.38	6415	24.55

#	Data	Tempo dalla mezzanotte (s)	Tempo dall' inizio (s)	RH_{RAW}	$RH(\%)$	T_{RAW}	$T(\%)$
41	2020 12 3	50423.00	16.018345	1345	44.38	6414	24.54
42	2020 12 3	50423.00	16.418462	1345	44.37	6415	24.55
43	2020 12 3	50424.00	16.818943	1345	44.38	6416	24.56
44	2020 12 3	50424.00	17.219302	1344	44.34	6418	24.58
45	2020 12 3	50424.00	17.619490	1344	44.35	6418	24.58
46	2020 12 3	50424.00	17.619490	1344	44.35	6418	24.58
47	2020 12 3	50425.00	18.019858	1345	44.38	6417	24.57
48	2020 12 3	50425.00	18.420379	1345	44.38	6418	24.58
49	2020 12 3	50426.00	18.820791	1344	44.35	6418	24.58
50	2020 12 3	50426.00	19.221040	1344	44.35	6416	24.56
51	2020 12 3	50426.00	19.621353	1344	44.34	6417	24.57
52	2020 12 3	50427.00	20.021686	1344	44.35	6418	24.58
53	2020 12 3	50427.00	20.422099	1344	44.35	6419	24.59
54	2020 12 3	50428.00	20.822469	1343	44.32	6418	24.58
55	2020 12 3	50428.00	21.222624	1344	44.35	6418	24.58
56	2020 12 3	50428.00	21.623074	1344	44.35	6419	24.59
57	2020 12 3	50429.00	22.023483	1344	44.35	6419	24.59
58	2020 12 3	50429.00	22.423907	1344	44.35	6420	24.60
59	2020 12 3	50430.00	22.824401	1343	44.32	6420	24.60
60	2020 12 3	50430.00	23.224921	1344	44.35	6420	24.60
61	2020 12 3	50430.00	23.625285	1344	44.35	6421	24.61
62	2020 12 3	50430.00	23.625285	1343	44.32	6421	24.61
63	2020 12 3	50431.00	24.025696	1343	44.32	6420	24.60
64	2020 12 3	50431.00	24.426193	1342	44.28	6421	24.61
65	2020 12 3	50432.00	24.826689	1342	44.29	6421	24.61
66	2020 12 3	50432.00	25.226887	1343	44.32	6422	24.62
67	2020 12 3	50432.00	25.627231	1342	44.29	6421	24.61
68	2020 12 3	50433.00	26.027469	1342	44.29	6420	24.60
69	2020 12 3	50433.00	26.427677	1342	44.28	6420	24.60
70	2020 12 3	50434.00	26.828121	1343	44.32	6420	24.60
71	2020 12 3	50434.00	27.228490	1342	44.28	6420	24.60
72	2020 12 3	50434.00	27.628931	1342	44.28	6420	24.60
73	2020 12 3	50435.00	28.029389	1343	44.32	6420	24.60
74	2020 12 3	50435.00	28.429974	1342	44.28	6420	24.60
75	2020 12 3	50436.00	28.830536	1342	44.28	6420	24.60
76	2020 12 3	50436.00	29.231185	1342	44.28	6418	24.58
77	2020 12 3	50436.00	29.231185	1342	44.28	6419	24.59
78	2020 12 3	50436.00	29.631409	1342	44.28	6420	24.60
79	2020 12 3	50437.00	30.031553	1342	44.28	6420	24.60
80	2020 12 3	50437.00	30.431854	1342	44.28	6419	24.59

#	Data	Tempo dalla mezzanotte (s)	Tempo dall' inizio (s)	RH_{RAW}	$RH(\%)$	T_{RAW}	$T(\%)$
81	2020 12 3	50438.00	30.832164	1342	44.28	6420	24.60
82	2020 12 3	50438.00	31.232556	1342	44.28	6420	24.60
83	2020 12 3	50438.00	31.632708	1342	44.28	6420	24.60
84	2020 12 3	50439.00	32.032864	1343	44.32	6420	24.60
85	2020 12 3	50439.00	32.433048	1342	44.28	6420	24.60
86	2020 12 3	50440.00	32.833236	1342	44.28	6422	24.62
87	2020 12 3	50440.00	33.233428	1342	44.29	6421	24.61
88	2020 12 3	50440.00	33.633556	1342	44.29	6421	24.61
89	2020 12 3	50441.00	34.033962	1342	44.29	6420	24.60
90	2020 12 3	50441.00	34.434319	1342	44.28	6422	24.62
91	2020 12 3	50442.00	34.834670	1342	44.29	6421	24.61
92	2020 12 3	50442.00	35.235140	1342	44.29	6420	24.60
93	2020 12 3	50442.00	35.235140	1342	44.28	6420	24.60
94	2020 12 3	50442.00	35.635390	1342	44.28	6420	24.60
95	2020 12 3	50443.00	36.035779	1342	44.28	6420	24.60
96	2020 12 3	50443.00	36.435923	1342	44.28	6420	24.60
97	2020 12 3	50444.00	36.836417	1342	44.28	6421	24.61
98	2020 12 3	50444.00	37.236875	1342	44.29	6421	24.61
99	2020 12 3	50444.00	37.637380	1342	44.29	6421	24.61
100	2020 12 3	50445.00	38.037534	1342	44.29	6421	24.61

Appendice B: Esercizio di Machine Learning

Si allega l'esercizio svolto di Machine Learning.