# Flusso dei dati

- Fasi dell'analisi
- Descrizione delle varie fasi
- Struttura dei dati
- Esempi

# Fasi dell'analisi

Procedure automatizzate

Linguaggi compilati

Procedure manuali

Linguaggi interpretati

Dati grezzi ( raw data )

Uso di costanti di calibrazione

Dati fisici ( punti, energie )

Uso di algoritmi

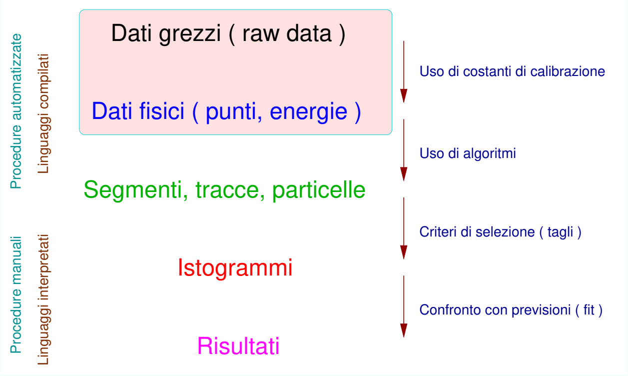
Segmenti, tracce, particelle

Criteri di selezione ( tagli )

Istogrammi

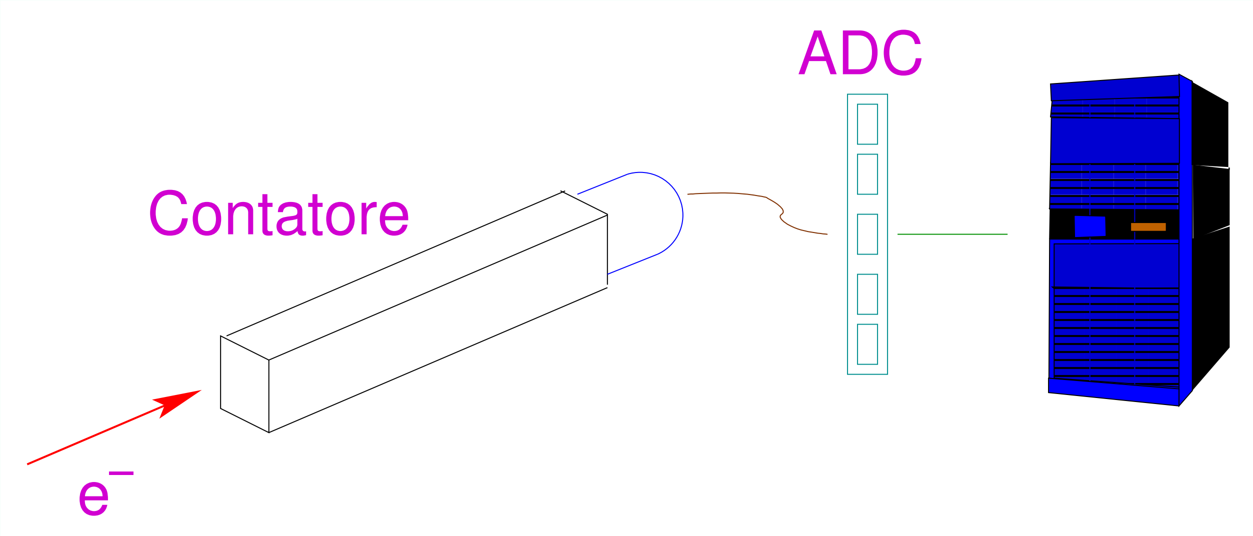
Confronto con previsioni ( fit )

Risultati



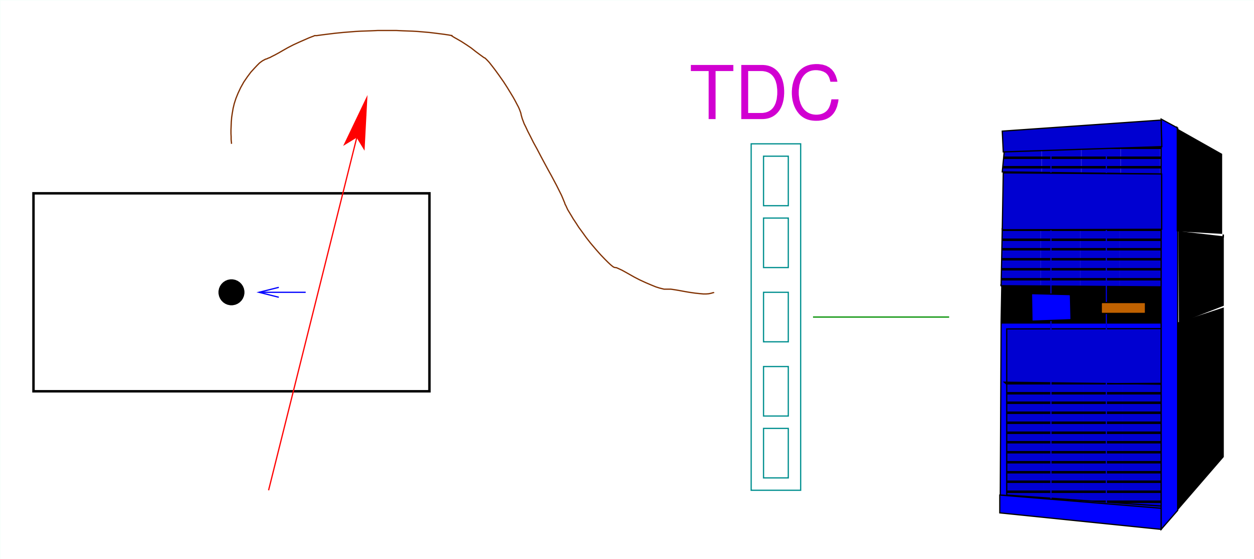

I "dati grezzi" sono costituiti da ciò che esce direttamente dall'apparato sperimentale: le grandezze fisiche vengono misurate da opportuni dispositivi digitali la cui uscita è quindi un numero che viene immagazzinato ( "data acquisition" ) . I parametri presenti nella relazione ( lineare nel caso più semplice ) che lega tale numero alla grandezza fisica vengono conservati in database a cui devono accedere i programmi che ricostruiscono le grandezze fisiche a partire dai dati. Poiché tale fase sta all'inizio di ogni processo di analisi e si ripete sempre in modo pressoché uguale i programmi che la svolgono vengono eseguiti in modo automatizzato ( "batch mode" ) con linguaggi compilati e quindi maggiormente efficienti. Per la grande mole di dati per questa fase si ricorre poi a "farm" di produzione, composte da numerosi processori.

# Esempio




Un elettrone attraversa un materiale denso e “sensibile” ( per es. piombo e scintillatore ) che emette luce, raccolta da un fotocatodo che emette una carica elettrica. La carica elettrica, proporzionale all’energia dell’elettrone, viene misurata da un convertitore analogico-digitale che fornisce un numero. Per riottenere l’energia tale numero ( che è il “dato grezzo” ) deve essere moltiplicato per un fattore ( “costante di calibrazione” ) determinato per ogni contatore.

# Esempio



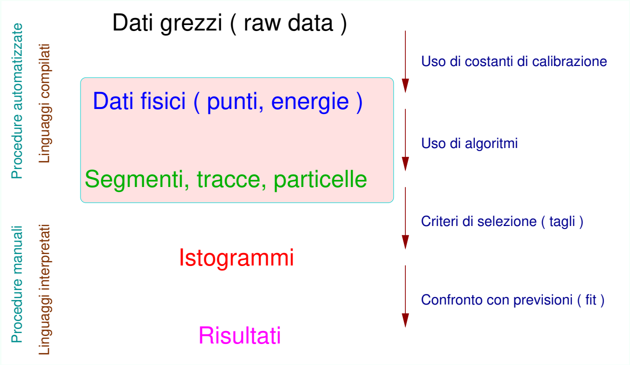

Una particella carica attraversa un tubo e ionizza il gas contenuto; gli elettroni prodotti raggiungono per effetto di un campo elettrico un anodo dove viene raccolto un segnale dopo un tempo proporzionale alla distanza della particella dall'anodo. Il "dato grezzo" è quindi costituito dal tempo, da cui si risale alla distanza mediante la velocità di deriva.

# Osservazioni

È opportuno osservare che la banalità delle operazioni mostrate è solo apparente: i dati grezzi sono usualmente scritti con un formato significativamente diverso per numero e posizione di bit da quello “nativo” dei processori e programmi usati per l’analisi. Oltre a ciò i vari canali di lettura sono identificati con indirizzi hardware di non immediata comprensione, che devono quindi essere tradotti secondo opportuni schemi di numerazione.

In un esperimento complesso vengono poi utilizzati numerosi apparati e dispositivi, i cui dati, raccolti contemporaneamente, vengono scritti insieme. Questo richiede che nell’analisi i dati provenienti da ciascun apparato vengano identificati ed estratti dai programmi, o parti di programma, dedicati alla loro elaborazione.

Tutte queste operazioni devono essere ripetute, in alcuni casi, per quantità enormi di dati, fino a parecchi Terabyte.

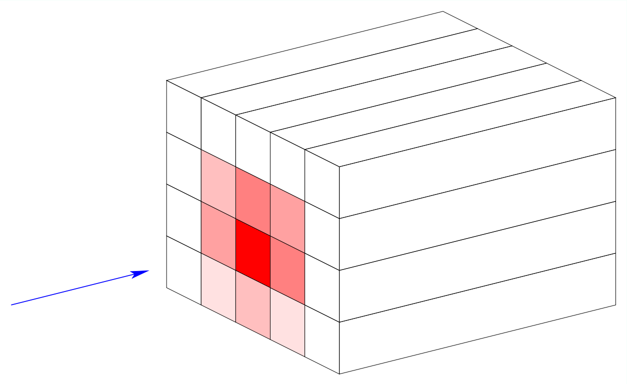



La semplice determinazione delle “grandezze fisiche” misurate dai singoli dispositivi è, ovviamente, nella gran parte dei casi ancora molto lontana dalla completa descrizione dei fenomeni studiati: la fase successiva consiste quindi nel combinare i dati tra loro per estrarre informazioni piú significative.

Questa fase non è usualmente costituita da un unico passo: inizialmente vengono combinate le misure provenienti da dispositivi simili costituenti un unico apparato ( “pattern recognition” ) ; i risultati cosí ottenuti vengono poi associati tra loro per ottenere una piú completa descrizione degli eventi.

Per eseguire questa operazione devono essere note le posizioni dei vari dispositivi, che vengono quindi anch’esse conservate in opportuni database.

# Esempio

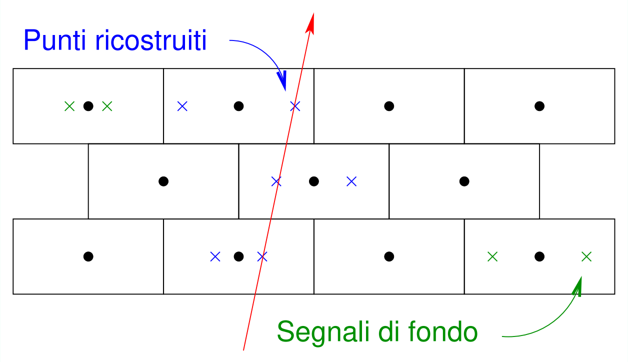


Ricostruzione dell'energia totale e della posizione di un elettrone:

- $\rightsquigarrow$ Ricerca del contatore con energia piú elevata
- $\rightsquigarrow$ Ricerca dei contatori adiacenti con energia maggiore di una soglia
- $\rightsquigarrow$ Somma delle energie $\Rightarrow$ energia totale
- $\rightsquigarrow$ Media pesata ( con le energie ) delle posizioni dei contatori colpiti $\Rightarrow$ posizione

La ricerca dei contatori vicini è apparentemente banale, ma in alcuni casi può diventare molto complicata!

# Esempio




Ricostruzione di una traiettoria ( traccia ) a partire dai punti misurati da un sistema di celle:

- Iterazione sui punti del primo piano
- Iterazione sui punti del piano adiacente
  - Iterazione su tutti i piani
- Costruzione di segmenti mediante interpolazione lineare
- Scelta dei segmenti “migliori”

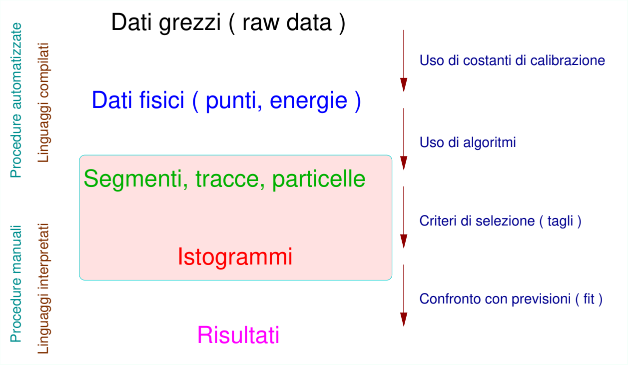
Il numero di possibili combinazioni può facilmente divergere: è necessario costruire algoritmi che limitino la scelta dei punti senza introdurre inefficienze.

# Ricostruzioni avanzate

La ricostruzione degli eventi studiati non necessariamente termina in una fase prefissata sempre uguale: possono infatti venire selezionati particolari campioni di dati per i quali è possibile approfondire la ricostruzione ( per esempio identificare le tracce lasciate dai prodotti del decadimento di un'unica particella ). D'altra parte può accadere che sia necessario ripetere alcune parti dell'analisi, per esempio per utilizzare parametri piú precisi, per correggere effetti sistematici o per utilizzare algoritmi dedicati.

Le varie fasi dell'analisi dei dati raccolti in un esperimento non sono quindi univocamente definite: i dati devono perciò venire organizzati in strutture che permettono a ogni livello di accedere ai dati di livello superiore.

Ciò veniva ottenuto, nelle analisi condotte con il linguaggio FORTRAN, mediante opportune librerie ( per es. ZEBRA ) che gestisono puntatori agli elementi di grandi vettori in cui sono conservati i dati. La programmazione ad oggetti consente invece di esprimere in modo più naturale le dipendenze esistenti tra dati di diversa natura.

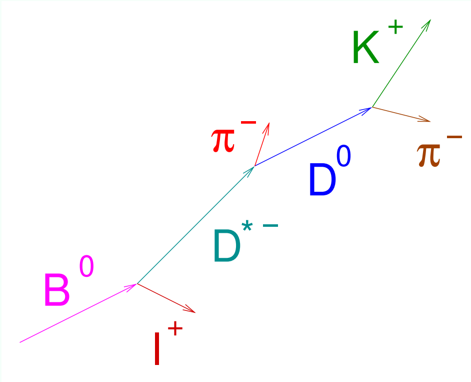



Una volta fatta la ricostruzione degli eventi studiati l'attenzione può concentrarsi sulle grandezze piú significative per la ricerca del risultato cercato; tali grandezze ( per esempio energie, angoli, molteplicità ) vengono quindi istogrammate.

Ovviamente nell'insieme di misure raccolte sono generalmente presenti dati spurii, chiamati "fondo", il cui contributo deve essere stimato, e possibilmente ridotto al minimo, per valutare correttamente i dati significativi, chiamati "segnale".

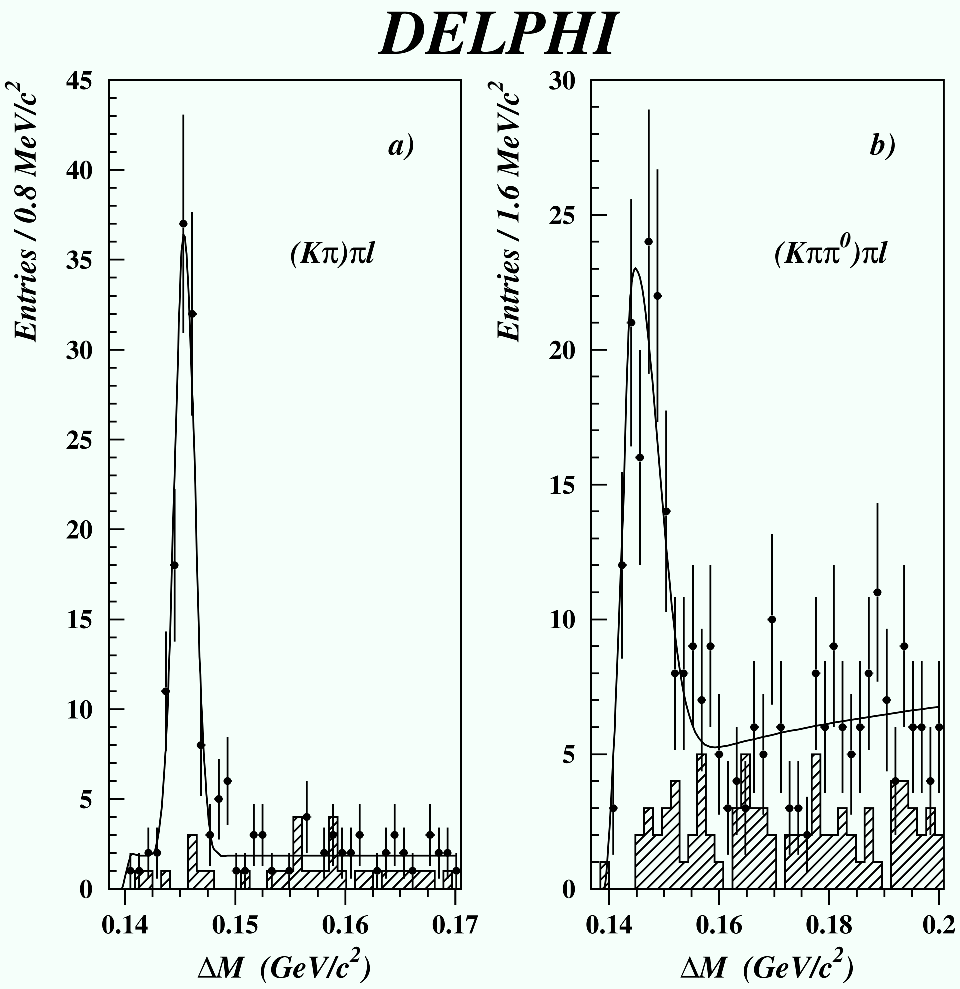
La separazione del segnale dal fondo, e in alcuni casi anche la determinazione del risultato stesso, viene usualmente condotta confrontando tra loro varie grandezze e studiando la distribuzione di alcune quando altre appartengono ( o non appartengono ) a certi intervalli.

# Esempio




Nel segnale si hanno K e l con carica uguale, nel fondo le cariche sono casuali.

$\Rightarrow$ Si costruiscono istogrammi di M($K\pi\pi$)-M($K\pi$) con le due combinazioni.

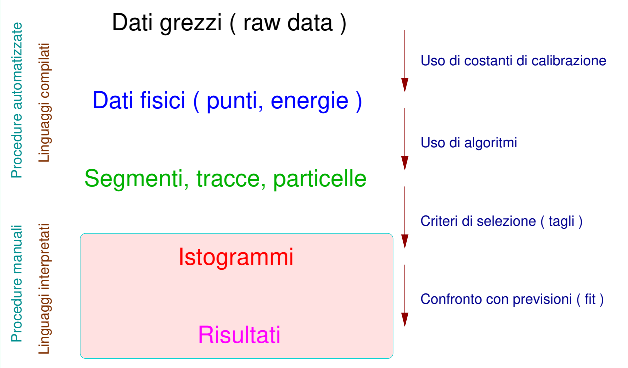


# N-tuple

La scelta delle variabili da correlare e la definizione dei criteri di selezione avviene, ovviamente, durante l'analisi stessa e non può essere fatta a priori; d'altra parte non è pensabile un metodo di analisi che preveda la ripetizione completa dell'analisi stessa ogni volta che viene modificato un valore di soglia o viene studiata la correlazione con una nuova variabile.

Viene comunemente chiamata n-tupla una raccolta di liste ordinate di variabili, che possono venire istogrammate singolarmente o in funzione l'una dell'altra: l'uso di n-tuple permette quindi di ripetere l'analisi di un insieme di variabili semplicemente rileggendo tali liste, senza partire ogni volta dai dati grezzi.

Costituiscono utilissimi strumenti per l'analisi dei dati alcuni programmi che leggono n-tuple scritte in formato definito e ne costruiscono istogrammi ed altre rappresentazioni, secondo criteri definiti interattivamente, permettendo di visualizzare graficamente i risultati.




Dagli istogrammi prodotti si arriva, nell'ultima fase, al risultato finale; ciò si ottiene usualmente confrontando con le previsioni le distribuzioni ottenute. La semplice esecuzione di calcoli algebrici, come la determinazione di medie o scarti quadratici, non è infatti quasi mai applicabile per molteplici motivi:

- le distribuzioni non sono gaussiane ( talvolta neppure esprimibili con formule analitiche )
- nelle distribuzioni sono presenti piú componenti che non si possono separare a livello del singolo dato
- alcune regioni delle distribuzioni possono essere inaffidabili per la presenza di disturbi e devono quindi essere escluse dall'analisi
- ...

# Distribuzioni di riferimento e fit

Il confronto tra distribuzioni sperimentali e previste viene eseguita esprimendo queste ultime in funzione di parametri corrispondenti alle grandezze fisiche la cui misura è l'obiettivo dell'analisi. Spesso, come detto, non è possibile esprimere tali distribuzioni in modo analitico; è necessario allora costruire delle distribuzioni di riferimento.

In pratica vengono costruiti istogrammi, analoghi a quelli ottenuti dai dati oggetto di studio, la cui forma esatta può essere variata agendo sui parametri che devono essere determinati. Questi possono corrispondere alla normalizzazione complessiva dell'istogramma di riferimento, o semplicemente ad una traslazione; in alcuni casi piú complicati permettono di calcolare dei fattori ( chiamati "pesi" ) da applicare ai vari canali dell'istogramma o anche ai singoli eventi istogrammati.

In questi casi il confronto tra istogrammi diventa puramente numerico; devono quindi essere identificate delle tecniche che permettano, su base numerica, di eseguire la determinazione dei parametri corrispondenti al migliore accordo.

# Esempio

La distribuzione studiata ( relativa ad una quantità di moto ) è rappresentata dai punti mentre gli istogrammi con diverso colore rappresentano varie componenti di cui viene determinata la normalizzazione. Oltre a ciò ogni evento, negli istogrammi sommati, viene pesato in base ad una variabile indipendente.