INTRODUZIONE OPERATIVA AL CONCETTO DEI VALORI CARATTERISTICI SECONDO L' EUROCODICE 7 (norma europea prEN1997-1)

Luca "McCoy" Nori – Geologo

(Versione - 28 marzo 2009)

La presente bozza è stata oggetto di revisione pubblica nel forum dei geologi, sezione geotecnica, al seguente link: http://www.geoforum.it/ubbthreads.php?ubb=newpost&Board=2
La versione definitiva è in preparazione.

Riassunto

L'Eurocodice 7: "*Eurocode 7: Geotechnical design - Part 1: General rules*" [1], introduce il concetto dei valori caratteristici dei parametri geotecnici. Il valore caratteristico, inteso come una stima cautelativa del parametro che influenza l'insorgere dello stato limite in considerazione, dovrà essere utilizzato in qualsiasi tipo di verifica geotecnica, che si tratti di SLU (stati limite ultimi ovvero potenziale presenza di una superficie di rottura) o di SLE (stati limite di esercizio ossia deformazioni di tipo elastico o di consolidazione a prescindere dallo stato di rottura). Lo stesso concetto fa parte della più ampia trattazione agli stati limite (SL), volta ad armonizzare la progettazione strutturale con quella geotecnica. Sono ormai trascorsi 5 anni dall'emanazione della versione definitiva degli Eurocodici e tra i professionisti che si dedicano alle verifiche geotecniche esistono ancora dubbi ed incertezze sulla determinazione dei valori caratteristici. Il DM 14/01/2008, "*Norme tecniche per le costruzioni*", ha adottato il concetto dei valori caratteristici, senza peraltro chiarirne la definizione e determinazione. La Circolare del Consiglio Superiore dei Lavori Pubblici, pubblicata nel sito della Regione Toscana [2], con carattere ancora ufficioso, chiarisce alcuni dettagli relativi alla determinazione dei valori caratteristici ma non offre metodi operativi per la loro determinazione. In letteratura tuttavia è presente materiale specifico [3,4,5,6,7,8,16,17,22], dal quale è possibile attingere per chiarire gli aspetti pratici di ordine generale inerenti all'applicazione dei valori caratteristici. Il presente lavoro vuole costituire una sintesi critica ed esegetica di utilità pratica, al momento non ancora esistente in lingua italiana. Sono inoltre presenti alcune proposte originali che verranno eventualmente approfondite in successivi lavori.

1. Definizione

L'EC7, al punto 2.4.5.2 2(P), definisce quanto segue:

"The characteristic value of a geotechnical parameter shall be selected as a cautious estimate of the value affecting the occurrence of the limit state"

Ossia :

"Il valore caratteristico di un parametro geotecnico sarà scelto come una stima cautelativa del valore che influenza l'insorgere dello stato limite"

I punti salienti della precedente definizione sono i seguenti:

- *Stima cautelativa*: si tratta di una *stima* (il vocabolo prende in considerazione l'incertezza esistente in geotecnica dovuta alla notevole variabilità delle proprietà dei depositi nonché all’incertezza dovuta alle informazioni non complete generalmente a disposizione tramite prove in situ e di laboratorio), che deve essere *a favore della sicurezza*.

- *Valore che influenza l’insorgere dello stato limite:* il valore caratteristico è in funzione dello stato limite considerato, ad esempio la rottura del terreno al collasso verticale della fondazione. Se esaminiamo, invece della rottura al collasso verticale, la rottura allo slittamento, il valore del parametro geotecnico sarà generalmente diverso.

Definire il valore caratteristico significa pertanto scegliere il parametro geotecnico che influenza il comportamento del terreno in quel determinato stato limite, ed adottarne un valore, o stima, a favore della sicurezza.

Si pone in evidenza che l’unica metodologia delineata dall’EC7 per la definizione dei valori caratteristici è di natura statistica. Questa non è resa obbligatoria (come vedremo non costituisce un ‘principio’ e pertanto non deve essere necessariamente adottata dai singoli stati membro); tuttavia, non vengono esplicitamente espressi altri metodi di natura oggettiva, se non, al punto 2.4.5.2 12(P), l’eventuale utilizzo di tavole standardizzate (a discrezione degli stati membro). Il ricorso a valori caratteristici tabulati ha un costo, che consiste nella maggiorazione della cautela ed il conseguente sovradimensionamento delle strutture di fondazione; infatti il citato punto precisa che “*When using standard tables of characteristic values related to soil investigation parameters, the characteristic value shall be selected as a <u>very cautious</u> value*” (la sottolineatura non è presente nel testo della normativa).

L’EC7, sempre nel paragrafo 2.4.5.2, enumera altri aspetti importanti nella scelta dei valori caratteristici, alcuni dei quali verranno trattati in seguito quando rilevante. Si desidera rammentare comunque che la stima cautelativa comporta generalmente valori più bassi del valore medio dei dati a disposizione, ma talora può comportare valori più alti, in particolari circostanze (ad esempio: attrito negativo lungo il fusto di pali, densità del terreno nelle verifiche di stabilità dei pendii, cedimenti differenziali)

2. Il metodo statistico

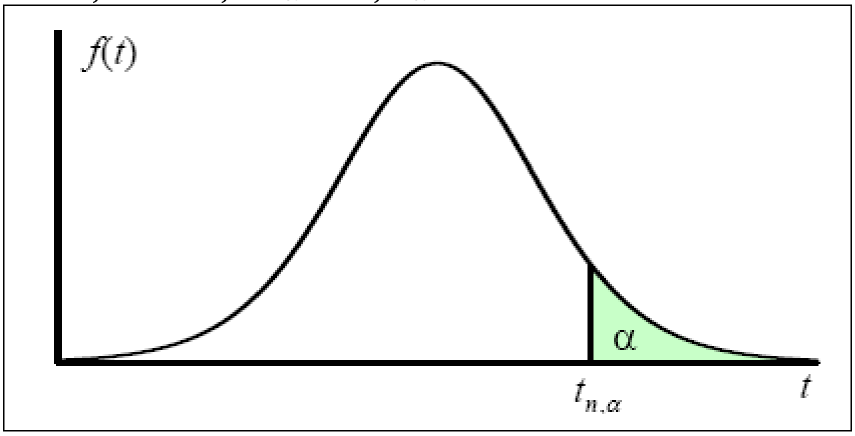
Il metodo statistico per la determinazione dei valori caratteristici non è accettato all’unanimità, soprattutto in ambito accademico. Si teme un ridimensionamento del giudizio tecnico e dell’esperienza professionale e un utilizzo meccanico della statistica; si sottolinea inoltre la frequente scarsità di dati per un corretto utilizzo della metodologia statistica. Tuttavia, sono ormai numerosi i lavori di letteratura che hanno dimostrato come gruppi numerosi di ingegneri geotecnici non sono in grado di pervenire ad un accordo nell’adozione di stime cautelative dei parametri geotecnici. Ad esempio, Bond e Harris [5] hanno sottoposto casi con trend crescente di NSPT e Cu con la profondità a oltre 100 ingegneri geotecnici inglesi, chiedendo di indicare quale fosse la stima cautelativa a loro giudizio. I risultati sono molto vari e in certi casi decisamente paradossali. Gli autori concludono come segue:
“*Una conclusione di questi studi è che gli ingegneri non sono particolarmente bravi nel selezionare una stima cautelativa del valore geotecnico di riferimento, specialmente quando i dati disponibili sono dispersi. Il trattamento statistico di campioni numerosi può guidare gli ingegneri in questo compito".*
Analogamente, Fellin [6] ha sottoposto un problema simile a 90 ingegneri geotecnici tedeschi, evidenziando la cospicua dispersione nella scelta dei valori caratteristici, che talora appaiono essere a sfavore, piuttosto che a favore, della sicurezza. Altri esempi del genere sono contenuti in Kruse [7] e Lacasse [8].

L’applicazione del metodo statistico è molto brevemente delineata nel punto 2.4.5.2 11 e la relativa nota:

“If statistical methods are used, the characteristic value should be derived such that the calculated probability of a worse value governing the occurrence of the limit state under consideration is not greater than 5%.
NOTE In this respect, a cautious estimate of the mean value is a selection of the mean value of the limited set of geotechnical parameter values, with a confidence level of 95%; where local failure is concerned, a cautious estimate of the low value is a 5% fractile”.

Se si utilizzano metodi statistici (non obbligatori come già notato), la derivazione del valore caratteristico deve essere tale che la probabilità calcolata di un valore peggiore (più sfavorevole) che governa l’insorgere dello stato limite in considerazione non sia maggiore del 5%. Si tratta pertanto di un margine conservativo del 5% (che può coincidere con un 5° percentile od un 95° percentile della distribuzione statistica in considerazione), il quale ci garantisce probabilisticamente di avere un 95% dei casi per i quali il valore caratteristico ci cautela. Il valore del 95% è anche quello indicato come probabilità ‘u’, o integrale della funzione, nelle tavole statistiche relative ai percentili della legge di Student [9,10] come illustrato in figura:

Figura 1: notazione per l’utilizzo dei percentili della variabile casuale *t* di Student; si noti che, con α = 0.05, 1-α = 0.95



A quest’esordio iniziale, di per sé non specifico, la nota aggiunge i due seguenti punti di importanza:

1. Una stima cautelativa del valore medio è una selezione del valore medio del limitato insieme dei valori del parametro geotecnico, con un livello di confidenza del 95%.
2. Dove l’analisi riguarda una rottura locale, una stima cautelativa del ‘low value’ è il 5° frattile (o percentile, termini qui utilizzati come sinonimi).

Data la formulazione incompleta e forse troppo sintetica dei precedenti punti, non meraviglia che il concetto di valore caratteristico si presti tuttora a dubbi e fraintesi. L’interpretazione corretta è tuttavia la seguente, come risulta evidente da autorevole letteratura tecnica, tra i quali autori è presente Niels Krebs Ovesen, presidente della commissione geotecnica degli Eurocodici [3]:

- **Precedente punto 1:** si cita il valore medio in quanto molto di frequente è questo il valore che viene ritenuto governare l’insorgere dello stato limite. Si tratta del valore medio del campione dei dati (campione generalmente piuttosto limitato in geotecnica) all’interno del volume di suolo (o dello spessore, lungo la profondità) interessato dalla potenziale superficie di rottura.

La frase una ‘selezione’ del valor medio indica che dobbiamo scegliere tra un set di valori medi. Quantunque la media del parametro all’interno del volume di suolo interessato dalla potenziale superficie di rottura sia unica, il fatto che noi estraiamo un campione generalmente limitato di dati implica l’esistenza di una incertezza, la cosiddetta incertezza epistemica. Ossia, se abbiamo estratto un campione con un Φ medio ad esempio di 36°, non è detto che un altro campione casualmente estratto nello stesso volume di suolo esibisca di nuovo una media di 36°. Ne discende che il calcolo da eseguire è quello dell’intervallo di confidenza della media al 90%, il cui lower (o upper bound) delimiti una regione con livello di confidenza del 95%. Quanto precede rammentando che se: Pr (x_1 < X < x_2) = 0.9, per motivi di simmetria Pr (X < x_2) = 0.95.
La citata frase del punto 1 corrisponde pertanto ad affermare che bisogna selezionare un 5° percentile della distribuzione della media. La distribuzione del valor medio, nel caso di distribuzione normale, viene generalmente calcolata utilizzando la ‘t’ di Student ad n-1 gradi di libertà ed ipotizzando che la media del campione coincida con la media della popolazione, usando la seguente relazione:

Equazione 1

$$\boxed{x_k = \bar{x} \pm t_{n-1}^{0.95}\left(\frac{s}{\sqrt{n-1}}\right)}$$

Dove:
x_k è il valore caratteristico desiderato
x con barra il valore medio (ignoto) della popolazione, ipotizzato essere uguale al valore medio del campione
t è il valore della distribuzione di student ad n-1 gradi di libertà con probabilità u = 95% (ossia, 1-α = 0.95 o alternativamente, α = 0.05)
s è la deviazione standard del campione
n il numero di dati.

Si pone in evidenza che in questo lavoro, a differenza di altri, si utilizza l’unbiased estimator per la deviazione standard della popolazione, per cui il dividendo sotto radice risulta essere n-1; l’alternativa più comune, ma generalmente meno cautelativa, è quella di utilizzare lo stimatore di massima verosimiglianza: si veda a proposito l’appendice A. Il valore di x_k in formula è agevolmente derivabile con foglio elettronico, essendo i parametri di input noti; la *t* di student per il percentile corrispondente a p=0.05 è individuabile ad esempio mediante la funzione Excel™ INV.T(2*p, n-1), nel nostro caso pari a INV.T(0.1, n-1). Al riguardo si pone in evidenza quanto chiarito nella guida in linea di Excel™:

- *La funzione INV.T viene calcolata come INV.T = p(t<X), dove X è una variabile casuale che segue la distribuzione t.*
- *È possibile restituire un valore t a una coda sostituendo probabilità con 2*probabilità. Il valore a due code di una probabilità 0.05 e 10 gradi di libertà si calcola con INV.T(0.05,10), che restituisce 2.28139. È possibile calcolare il valore a una coda di una probabilità e gradi di libertà uguali con INV.T(2*0.05,10), che restituisce 1.812462.*

Per altri dettagli relativi agli intervalli di confidenza esistono vari lavori nel pubblico dominio, uno dei quali a titolo di esempio è citato in bibliografia [11]

- **Precedente punto 2**: si differenzia dall’antecedente nel caso in cui la rottura sia ‘locale’, ossia interessi volumi relativamente piccoli del terreno. Questo avviene generalmente in casi di verifica alla rottura alla punta di pali di piccolo diametro e micropali o di fondazioni di larghezza contenuta, ma è comunque da verificare in base alle oscillazioni stocastiche dei valori di resistenza del terreno (tali oscillazioni vengono materializzate ad esempio dal segnale delle penetrometrie continue). Se la superficie di rottura può svilupparsi in un piccolo volume di terreno, lo stato limite ultimo potrebbe essere governato da una fluttuazione statistica meno resistente e non dalla media di tutte le fluttuazioni, come avviene nel caso di grandi volumi di rottura. In definitiva, nell'ipotesi di piccoli volumi interessati dalla superficie di rottura, il valore caratteristico consiste nel 5° percentile della distribuzione del campione (e non della media campionaria). Questa è la interpretazione più comune del valore caratteristico da parte della comunità dei tecnici; tale interpretazione risulta incompleta ed erronea se applicata indiscriminatamente, ossia non in relazione al volume interessato dalla rottura. La formula da applicare pertanto è la seguente:

Equazione 2

$$\boxed{x_k = \bar{x} \pm z_{0.05} \cdot s \approx \bar{x} \pm 1.645 \cdot s}$$

Dove:
x_k è il valore caratteristico desiderato
x con barra il valore medio (ignoto) della popolazione, ipotizzato essere uguale al valore medio del campione
z è la distribuzione normale standardizzata
s è la deviazione standard del campione

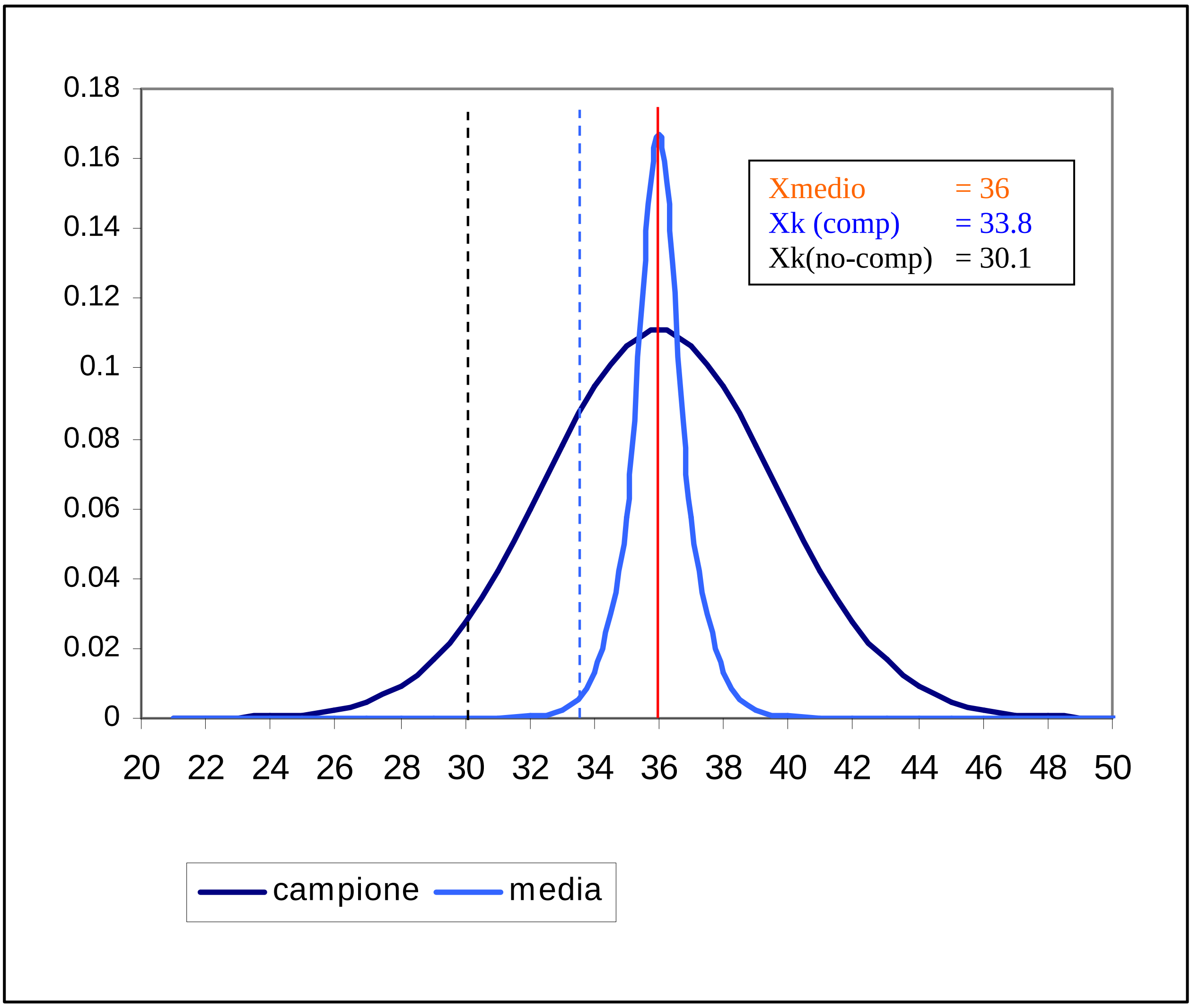
In questo caso, per ottenere direttamente il valore caratteristico x_k, è conveniente ricorrere ad esempio alla funzione Excel™ INV.NORM(0.05,media,deviazione standard), che ci fornisce direttamente il 5° percentile della distribuzione normale derivata dai parametri (media e deviazione campionarie) ricavati dai dati a disposizione.

A proposito dell'aspetto legato alla grandezza del volume di rottura, l'EC7 nel punto 2.4.5.2 (4)P esplicita:
"*The selection of characteristic values for geotechnical parameters shall take account of the following:… the extent of the zone of ground governing the behaviour of the geotechnical structure at the limit state being considered*".

L'ampiezza della zona del terreno che governa il comportamento del sistema geotecnico nello stato limite considerato è pertanto uno dei fattori che incidono sulla scelta del valore caratteristico. Questo fattore, troppo spesso non adeguatamente considerato, è stato introdotto in quanto l'EC7 ha inteso inserire, semplificandoli, i concetti tipici della moderna trattazione geotecnica basata sulla geostatistica e sui campi aleatori; molto sinteticamente, ad ogni punto del terreno corrisponde una variabile aleatoria che caratterizza il parametro geotecnico in considerazione. Pertanto il segnale di una penetrometria dinamica lungo una determinata verticale, ad esempio, viene considerato come una realizzazione, o traiettoria, di un processo aleatorio, o segnale stocastico, o serie temporale, o serie spaziale. Noto il processo aleatorio in una successione omogenea, caratterizzato da media, varianza e autocovarianza, è noto il comportamento del terreno anche in relazione alle ciclicità tipiche dei parametri geotecnici. Tornando all'osservazione che spesso la verifica è governata dalla resistenza media lungo la superficie di rottura, come scrivono Nadim e Lacasse [12]: *"Maggiori sono le dimensioni entro le quali i parametri sono mediati, più le fluttuazioni attorno al trend tendono ad elidersi nel calcolo della media spaziale"*.

Alcuni approfondimenti sui processi aleatori sono disponibili nei links in bibliografia [13, 14, 15, 16].
Altro ulteriore e importante fattore da considerare è quello illustrato nel sopra citato punto 2.4.5.2 (4)P: "*The selection of characteristic values for geotechnical parameters shall take account of the following:… the ability of the geotechnical structure to transfer loads from weak to strong zones in the ground*". Che pone in evidenza il fatto che, se la struttura di fondazione è sufficientemente rigida, i carichi vengono trasferiti dalle zone deboli a quelle forti del terreno, con un meccanismo di compensazione strutturale simile al descritto meccanismo di compensazione spaziale delle fluttuazioni casuali dei parametri statistici. Questo concetto implica che, in presenza di strutture di fondazioni rigide, il valore caratteristico corrisponde al 5° percentile della media (equazione 1), anche se non si è in presenza di grandi volumi di rottura.

Figura 2: angolo di attrito laterale, in blu la distribuzione della media, in nero la distribuzione del campione. Il 5° percentile della distribuzione della media costituisce il valore caratteristico in presenza di compensazione ed è indicato con linea verticale tratteggiata blu; il suo valore è pari a 33.8. Il 5° percentile della distribuzione del campione costituisce il valore caratteristico in assenza di compensazione ed è indicato con linea verticale tratteggiata nera; il suo valore è pari a 30.1. La costruzione è valida esclusivamente per un campione di 10 dati con media = 36, varianza = 3.6^2. Tutti i valori sono in unità omogenee (gradi).



In sintesi, all’interno di uno strato omogeneo e dello spessore di influenza dello stato limite considerato, valgono le seguenti regole:

A. Se esiste compensazione spaziale (volume interessato dallo stato limite, o suo spessore in una dimensione, significativamente maggiore della lunghezza delle fluttuazioni nei parametri di resistenza del terreno), allora il valore caratteristico corrisponde al 5° percentile della media (equazione 1) ed è generalmente non molto distante dalla media stessa.
B. Se esiste compensazione strutturale (fondazioni sufficientemente rigide tali da distribuire le sollecitazioni omogeneamente sul terreno), allora il valore caratteristico è lo stesso del precedente caso A.
C. Se non esiste compensazione strutturale, né compensazione spaziale, allora il valore caratteristico è il 5° percentile della distribuzione statistica del campione, ed è generalmente piuttosto distante dalla media.

In figura 2 sono illustrati i due valori caratteristici che possono essere ottenuti dallo stesso campione di dati, relativi all’angolo di attrito di picco di uno strato omogeneo con media = 36° e deviazione standard = 3.6; la numerosità del campione = 10.

Si pone di nuovo in evidenza che nella maggior parte della letteratura tecnica specifica il denominatore dell’equazione 1 viene presentato come un ‘n’ sotto radice, il che comporta valori leggermente superiori del valore caratteristico, in funzione di n stesso (nell’esempio di figura 1, x_k con compensazione avrebbe assunto il valore di 33.9 invece di 33.8). Nel presente lavoro si ritiene maggiormente corretto ricorrere al denominatore sotto radice ‘n-1’, per i motivi in precedenza citati e sintetizzati nell’’appendice A.

3. Procedura operativa

Trascurando per adesso i casi particolari (quantunque non infrequenti) di distribuzioni diverse dalla normale e di presenza di un trend nei dati, casi esaminati in successivi capitoli, è possibile delineare la seguente lista ordinata di operazioni da svolgere per arrivare alla stima finale del valore caratteristico:

1. Trasformare i valori della prova nel parametro di interesse (non applicato a prove di laboratorio)
2. Trasformare ulteriormente i parametri geotecnici per adattarli alla verifica da svolgere (ad esempio, trasformazione del Φ triassiale in Φ in condizioni piane, oppure trasformazione della Cu da vane test con la correlazione di Bjerrum in funzione di IP)
3. Trasformare eventualmente i parametri in condizioni statiche a parametri in condizioni sismiche
4. Individuare il volume, o spessore, di influenza per lo SLU o SLE da verificare
5. Individuare i dati presenti in tale spessore (i dati possono provenire da tutte le prove a disposizione se lo strato è omogeneo)
6. Verificare se esiste la presenza di compensazione strutturale (sistema di fondazione rigido)
7. Se il punto 6 è verificato, applicare l’equazione (1) per trovare il valore caratteristico, che generalmente sarà ubicato a non grande distanza dal valore medio del parametro
8. Se il punto 6 non è verificato, effettuare uno studio del segnale per accertare se esista o meno compensazione spaziale, ossia se il volume (o spessore) interessato dallo SL sia grande o piccolo in relazione alle fluttuazioni stocastiche del segnale
9. Se esiste compensazione spaziale, applicare l’equazione 1 per trovare il valore caratteristico, che generalmente sarà ubicato a non grande distanza dal valore medio del parametro

10. Se non esiste compensazione spaziale, applicare l'equazione 2 per trovare il valore caratteristico, che generalmente sarà ubicato a maggiore distanza dal valore medio del parametro rispetto al caso di mancanza di compensazione

Si pongono in evidenza alcuni aspetti legati alla procedura operativa delineata:

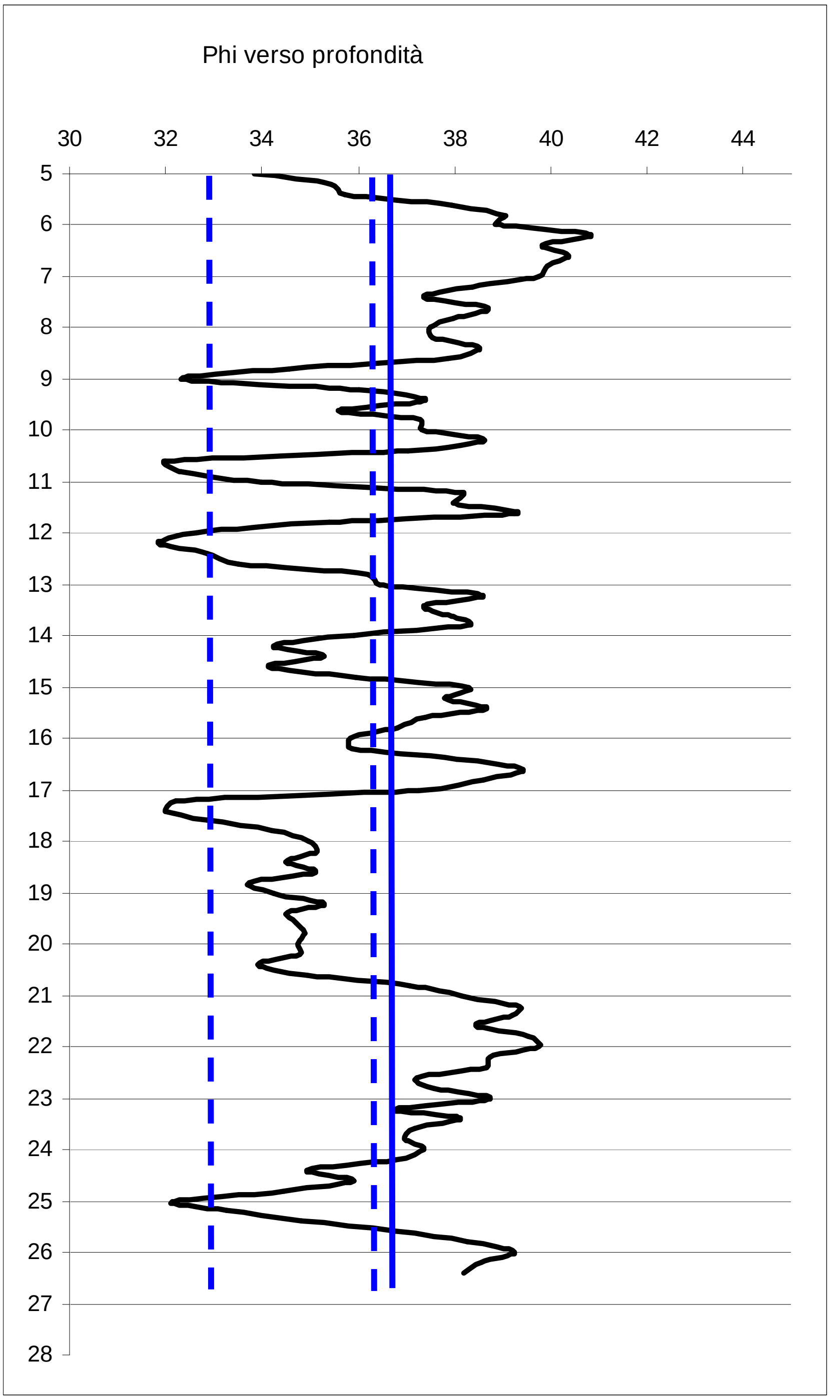
- Qualora la fondazione insista esattamente lungo la verticale di una delle prove, e se ne conoscano a priori la geometria e la profondità di incasso, allora la traiettoria del processo è nota e l'equazione da utilizzare è sempre la 1, con i dati ubicati all'interno dello spessore di influenza (la topologia delle gole e dei picchi del segnale è, per quella fondazione, nota)
- Se la verifica non comporta la trasformazione dei valori ottenuti dalla prova in parametri geotecnici, allora si tratteranno direttamente i valori della prova; alcuni esempi sono dati dalle formule che ricavano la capacità portante statica direttamente dai valori letti dal pressiometro, ad esempio, o dai valori di prove di tipo SPT o CPT, o dalla formula di Burland e Burbidge per calcolare i cedimenti direttamente dal valore di SPT.

In figura 3 è illustrato un tratto di segnale derivante da una penetrometria statica meccanica; sono evidenziati il valore medio, il valore caratteristico $x_{k,comp}$ e il valore caratteristico $x_{k,no\text{-}comp}$ per lo strato omogeneo.

Quello illustrato in figura è un esempio nel quale non è semplice accertare quando il volume che interessa la superficie di rottura sia piccolo o grande. Nel caso di un palo di fondazione, tuttavia, la resistenza lungo il fusto si svilupperebbe lungo uno spessore elevato, per cui il Φ_k = 36.2, di poco minore del Φ medio di 36.5; la resistenza lungo la punta sarebbe governata da un volume significativamente minore, probabilmente piccolo relativamente alle fluttuazioni del valore di Φ di minore frequenza, pertanto Φ_k = 32.9, significativamente minore del Φ medio di 36.5, a meno il palo faccia parte di una struttura di fondazione nel complesso rigida. Si pone in evidenza che il valore di $\Phi_{k,no\text{-}comp}$ = 36.2 è nel complesso rappresentativo della media dei valori di Φ relativi alle fluttuazioni di minore resistenza.

Alcune tipologie di indagini, come le prove penetrometriche SPT o le prove di laboratorio eseguite su carote estratte da sondaggi, non permettono di ricostruire la ciclicità del processo aleatorio; in tal caso la cautela suggerisce di procedere come se non esistesse compensazione spaziale. Altra soluzione potrebbe essere quella di ipotizzare la probabile presenza di compensazione solo per fondazioni di grande volume.

Figura 3: Strato sabbioso con Φ medio = 36.5, Φ_k in situazione di compensazione strutturale o spaziale = 36.2 , Φ_k non in situazione di compensazione strutturale o spaziale = 32.9 . Il segnale del penetrometro è in nero, la media in blu continuo, i valori caratteristici in blu tratteggiato. La numerosità del dataset n = 108



4. Campioni di scarsa numerosità

In geotecnica è circostanza decisamente non infrequente il dovere eseguire delle verifiche in presenza di scarsità di dati. Ciò accade generalmente in progetti di modesto rilievo, con budget di spesa di solito limitato, o quando si eseguono sondaggi con SPT o prove di laboratorio su campioni estratti, dei quali solo pochi interessano lo strato omogeneo o lo spessore di influenza dello SL. Una obiezione frequente al trattamento statistico è che questo non può essere eseguito con pochi dati. Al contrario, utilizzando la discriminazione e il giudizio tecnico e la conoscenza a priori regionale e locale, un trattamento statistico è possibile anche nel caso estremo di un solo dato a disposizione (o, al limite, di nessun dato disponibile, in questo caso facendo affidamento esclusivamente sulla conoscenza pregressa). Quando ci troviamo in condizioni di compensazione (generalmente strutturale) e il dataset ha una numerosità piccola (da 1 a 5 dati usualmente), possiamo ricorrere alla statistica classica, con ipotesi di varianza nota. L'equazione 1 assume pertanto la seguente forma:

Equazione 3

$$\boxed{x_k = \bar{x} \pm z_{0.05}\left(\frac{\sigma}{\sqrt{n}}\right) \approx \bar{x} \pm 1.645 \cdot \left(\frac{\sigma}{\sqrt{n}}\right)}$$

Dove:
x_k è il valore caratteristico desiderato
x con barra il valore medio (ignoto) della popolazione, ipotizzato essere uguale al valore medio del campione
z è la distribuzione normale standardizzata
σ è la deviazione standard della popolazione
n è la numerosità del campione

Si pone in evidenza quanto segue:

- Se $n \rightarrow \infty$ allora $t \rightarrow z$ e pertanto l'equazione 1 tende all'equazione 3; già da n =30 i valori di t e z sono a tutti gli effetti pratici coincidenti
- Con n = 1 (un singolo dato) il termine sotto radice diventa pari all'unità, pertanto l'equazione 3 tende all'equazione 2; ciò significa che, con un solo dato, i valori caratteristici in presenza e in assenza di compensazione coincidono
- L'ipotesi di valore del dato coincidente con la media del parametro nello strato in considerazione deve essere verificata tramite il giudizio tecnico e l'esperienza pregressa, o conoscenza a priori in quella determinata tipologia di depositi. Talora il controllo, ad esempio in un sondaggio, può essere effettuato tramite l'ispezione delle carote prelevate attorno al campione sottoposto a prove di laboratorio, eventualmente utilizzando come ausilio strumenti di affidabilità non elevata quali pocket penetrometer e pocket vane test.

Il valore di **σ** nell'equazione 3 andrebbe determinato con l'ausilio di database locali specifici; quando questi mancano, si possono utilizzare i valori reperibili in letteratura. A tale proposito è utile esprimere la variabilità in termini di coefficiente di variazione COV (anche V o CV), espressa anche in termini percentuali:

Equazione 4

$$\boxed{COV = \frac{\sigma}{\mu}} \qquad (4)$$

Dove:
σ = deviazione standard della popolazione
μ = media della popolazione

Schneider [17] al proposito raccomanda i seguenti valori di COV :

Angolo di attrito Φ:	10 %
Coesione:	40 %
Modulo di comprimibilità:	40 %

Lo stesso Schneider offre la seguente formula semplificata:

Equazione 5

$$\boxed{x_k = \bar{x} \cdot \left(1 - \frac{COV}{2}\right)}$$

divenuta piuttosto popolare, che però non comporta nessun reale vantaggio rispetto alle formule rigorose indicate in precedenza, le quali sono facilmente implementabili in fogli di calcolo,come già illustrato. La formula è simile all'equazione 3 quando $n \approx [10, 11]$ e pertanto può sottostimare il parametro con $n > 11$ e sovrastimarlo (a sfavore della sicurezza) se $n < 10$. Schneider a proposito si limita ad affermare che la relazione è stata utilizzata in Svizzera a partire dal 1990 con buoni risultati; in questo lavoro, tuttavia, se ne sconsiglia l'utilizzo per scopi rigorosi. L'equazione potrebbe invece essere vantaggiosamente utilizzata in situazioni di maggiore onere computazionale, quale il calcolo dell'intervallo di confidenza di una regressione lineare, come si vedrà più oltre.
Il trattamento statistico del singolo dato presuppone un discernimento tecnico nello stabilire se il dato stesso può essere ritenuto rappresentativo della media della popolazione nel deposito in esame. Tale discernimento può utilizzare conoscenza pregressa locale, regionale e altri indizi di natura geologica, sedimentologica e geotecnica.

5. Presenza di trend

Il processo aleatorio, visualizzato ad esempio dalla famiglia dei segnali delle penetrometrie eseguite in un sito, può essere caratterizzato da non stazionarietà, ossia può presentare un trend, crescente o decrescente con l'aumentare della profondità. In questo caso è possibile procedere in due modi:

1. Metodo approssimato: suddivisione del trend in intervalli, ognuno dei quali caratterizzato da propria media e valori caratteristici
2. Metodo rigoroso: calcolo dell'intervallo di confidenza e dei livelli di tolleranza sulla curva di regressione.

Entrambi i metodi sono presentati nelle figure 4, 5 e 6. In figura 4 abbiamo 2 realizzazioni, o traiettorie, dello stesso processo aleatorio non stazionario (segnali trasformati di 2 CPT meccaniche all'interno di una successione limoso-argillosa), rappresentante la diminuzione della Cu con l'incrementare della profondità. In figura 5 una funzione a gradini, visualizzata in rosso, approssima il trend decrescente. In figura 6 invece il trend è approssimato da una retta, con i suoi intervalli di confidenza al 90% bilaterali (intervalli della media e dei dati). Adottando il modello di figura 5 suddividiamo la successione in 3 spessori 'omogenei', ognuno dei quali esibisce i propri valori caratteristici. Con il modello di figura 6, invece, i valori caratteristici in presenza di compensazione sono quelli situati lungo i rami di iperboli che rappresentano i lower e upper bound dell'intervallo di confidenza al 90%, mentre i valori caratteristici in presenza di compensazione sono quelli situati lungo le rette più lontane da quelle di regressione , che rappresentano i livelli di confidenza al 5° e al 95° percentile rispettivamente (il 90% dei dati è in teoria compreso tra queste due rette).

Figura 4

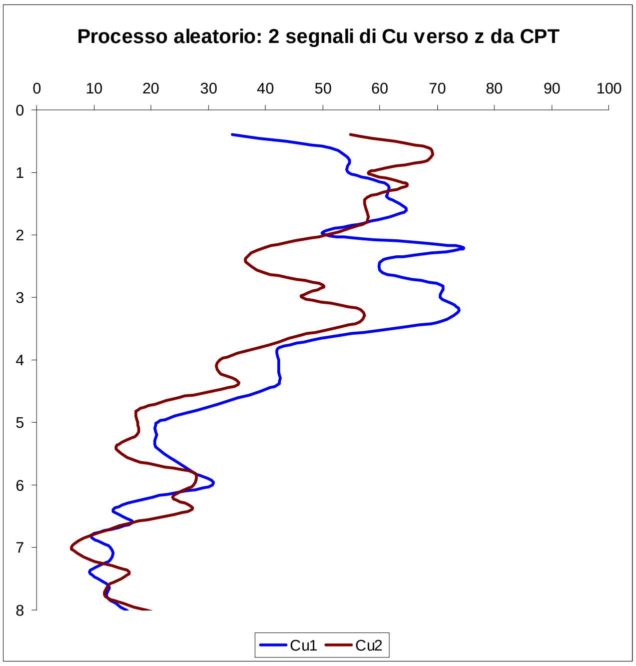


Figura 5

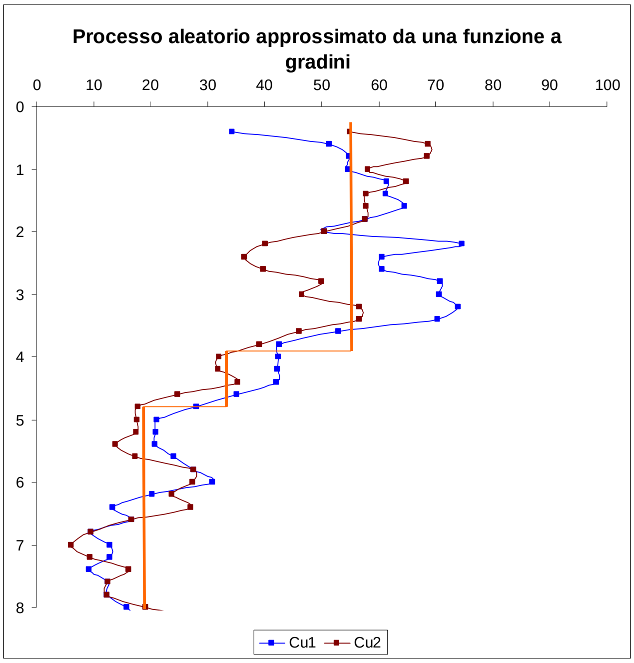
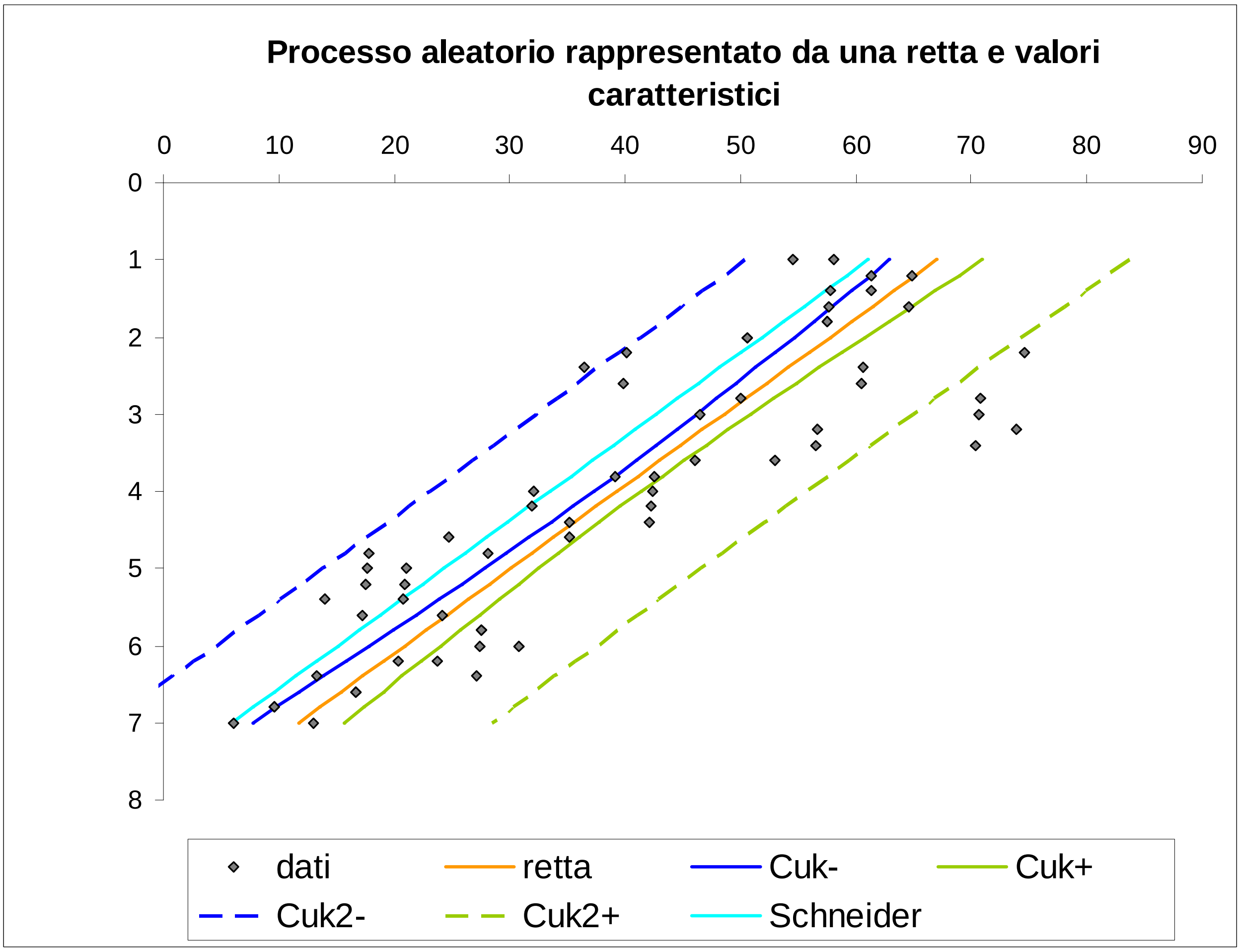


Figura 6



La procedura per ottenere le iperboli e le rette di previsione è descritta nei riferimenti [3,5,22] ed è riportata in appendice B. La procedura delineata andrebbe modificata al caso specifico di una regressione di X su z (profondità).

Tale procedura può essere facilmente implementata in un foglio elettronico; tuttavia, poiché l'applicazione può risultare leggermente laboriosa, è possibile ottenere dei valori approssimati come segue:

- Metodo di Schneider per i valori caratteristici in presenza di compensazione: il valore caratteristico dista di **s/2** dalla retta di regressione, dove s è la deviazione standard calcolata sui residui. La retta che descrive l'approssimazione di Schneider è in celeste nel grafico di figura 6.
- Metodo grafico approssimato per il 5° e 95° percentile: si tracciano due rette parallele a quella di regressione e da questa equidistanti, all'interno delle quali si venga a trovare il 90 % dei dati.

L'equazione 5 proposta da Schneider andrebbe applicata ai residui della regressione, ossia al dataset detrendizzato (reso stazionario). Nella pratica è sufficiente adottare la seguente equazione di correzione:

Equazione 6

$$\boxed{\varepsilon_x = X_z - \hat{x}_z}$$

Dove:

ε_x = residuo
X_z = valore del parametro geotecnico alla profondità z
$\hat{x}_z$ = valore del parametro geotecnico predetto dalla retta di regressione alla profondità z

Si pone inoltre in evidenza quanto segue:

- Si tratta di una regressione di X (parametro geotecnico) su z (profondità)
- La condizione di stazionarietà implica anche l'esistenza di omoschedasticità (varianza costante).

La deviazione standard dei residui calcolati applicando l'equazione 6 al caso di figura 6 permette di costruire la stima di Schneider del lower bound dell'intervallo di confidenza (retta celeste in figura 6). Tale retta caratteristica risulta essere chiaramente più cautelativa della curva iperbolica caratteristica (in blu) costruita secondo la procedura rigorosa.

Si pone in evidenza che Excel™ include, negli strumenti di analisi dati, la possibilità di effettuare modelli di regressione con graficizzazione automatica dei residui.

6. Distribuzioni diverse dalla normale

Le considerazioni esposte nel precedente capitolo sono rigorosamente valide ipotizzando che il parametro geotecnico appartenga ad un modello di distribuzione statistica normale caratterizzato da media e deviazione standard, ossia $X \sim N(\mu, \sigma)$. Tale modello è genericamente applicabile nella pratica per la distribuzione di Φ, caratterizzata solitamente da basse varianze. Un aspetto problematico dell'applicazione del modello normale in terreni puramente attritivi è la mancanza del vincoli inferiore sul valore minimo della distribuzione, Φ_{cv}, che è in realtà è > 0 e non piccolo. Ciò può portare a valori irrealistici dello stesso. Il problema può essere risolto adottando una distribuzione normale troncata o meglio un'altra distribuzione del tipo beta o PERT. Schneider [17] consiglia un caso specifico di tale distribuzione, valido tuttavia esclusivamente con l'esistenza di vincoli (l'autore non specifica tale aspetto). La possibilità di utilizzare una distribuzione PERT senza vincoli per il parametro Φ, con l'aiuto di un semplice foglio di calcolo, verrà esaminata in un successivo lavoro. Altra tipologia di distribuzione utilizzata soprattutto per la Cu ed E (modulo elastico o edometrico) è la distribuzione lognormale. Questa rappresenta in maniera piuttosto realistica i dati, con il problema dovuto tuttavia alla scelta del valore caratteristico con compensazione: $X_{k,comp}$; la distribuzione della media nella lognormale non è infatti derivabile analiticamente e una semplice trasformazione delle variabili risulta in una distribuzione della mediana e non della media. In statistica applicata si utilizzano il metodo esatto di Land, oppure metodi numerici quale il bootstrap o metodi approssimati quali quello di Angus e quello di Cox, illustrati nelle seguenti equazioni:

Equazione 7: metodo di angus

$$L_{1-\alpha}(\bar{Y}, S^2) = \bar{Y} + \frac{S^2}{2} - \frac{t_{1-\alpha/2}(n-1)}{\sqrt{n}} * \sqrt{S^2(1+\frac{S^2}{2}}$$

Equazione 8: metodo di Cox

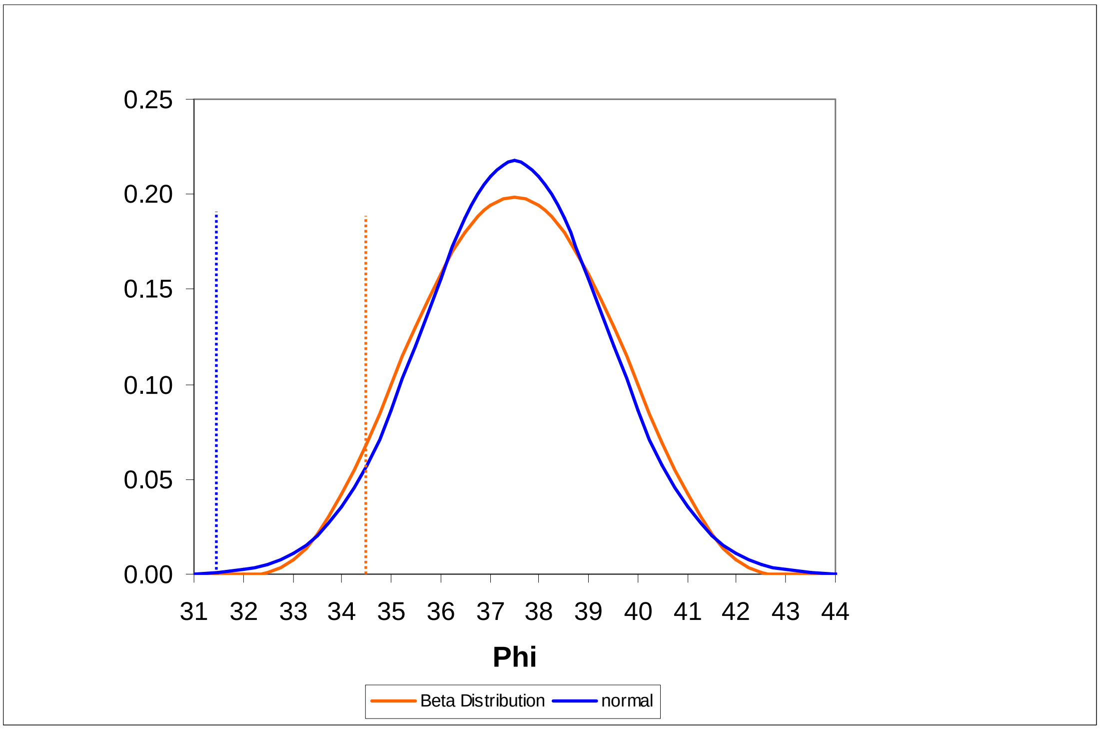
$$L_{1-\alpha}(\bar{Y}, S^2) = \bar{Y} + \frac{S^2}{2} \pm z_{1-\alpha/2} * \sqrt{\frac{S^2}{n} + \frac{S^4}{2(n-1)}}$$

Dove:

$\bar{Y}$ è il valore medio del dataset con distribuzione lognormale
z è la distribuzione normale standardizzata
t è il valore della distribuzione di student ad n-1 gradi di libertà
S è la deviazione standard del campione
n il numero di dati.

Ritornando alla distribuzione di Φ, è possibile utilizzare una distribuzione del gruppo beta per modellare con rigore il comportamento fisico dei terreni granulari. Il minimo della distribuzione coincide infatti con il Φ_{cv}, proprietà dello specifico terreno che non dipende dalle condizioni di sollecitazione ed addensamento, mentre il massimo può essere calcolato in funzione della densità relativa Dr, del coefficiente di uniformità CU e di alcuni parametri sperimentali delineati dalla letteratura tecnica [19]. Un esempio di distribuzione Beta-PERT è illustrato in figura 7, in rosso; tale aspetto verrà esposto in maggiore dettaglio in un prossimo lavoro.
Per adesso si intende porre in evidenza come in una sabbia uniforme con Φ_{cv} = 32° e Φ_{max}= 43°, Φ misurato ≡ Φ modale = 37.5°, la distribuzione beta-PERT risulta in un Φ_k = 34.5°, mentre il modello di distribuzione normale risulta in un Φ_k = 31.4°, significativamente più basso. Si noti anche che quest'ultimo valore risulti inferiore al valore del Φ_{cv}, condizione non realistica poiché il 5° percentile di una variabile aleatoria è sempre maggiore del minimo valore che può assumere tale variabile.

Figura 7: modello normale e beta-PERT nello stesso deposito di sabbia quarzoso-feldspatica; sono indicati i valori caratteristici per entrambi le distribuzioni



7. Prove di laboratorio con parametri efficaci

Avendo una o più coppie di parametri Φ - c derivanti da prove di laboratorio, è possibile determinare i valori caratteristici esattamente come esposto nei precedenti paragrafi. Questo se ipotizziamo l'indipendenza statistica tra i parametri. Nella realtà, è noto che Φ e c esibiscono generalmente une correlazione negativa, ossia all'incrementare di Φ si nota una diminuzione di c e viceversa. L'errore che si commette se si trascura tale correlazione non è grande ed è a favore della cautela. Tuttavia, è possibile adottare il metodo rigoroso se si è a conoscenza della correlazione tra i parametri; il valore del coefficiente ρ generalmente utilizzato in letteratura è quello proposto da Harr [20], elaborato da varie proposte di letteratura e pari a **-0.47**; altrimenti è necessario basarsi su dati più specifici, ad esempio quelli ricavati da Cherubini [21] sulle argille blu di Matera, nelle quali **ρ= -0.61**. Van Alboom e Mengè [22] consigliano una rappresentazione grafica che consiste nel tracciare la retta di regressione dei punti di rottura dei provini di laboratorio nel piano σ, τ. Alla regressione di τ su σ vengono quindi associati gli intervalli di confidenza e i livelli predittivi come già illustrato nel paragrafo 5 (presenza di trend). Tale rappresentazione ha lo svantaggio di richiedere un certo numero di provini e necessitare la linearizzazione del ramo di iperbole che rappresenta il lower bound (o alternativamente l'upper bound) dell'intervallo di confidenza nell'intervallo di sollecitazioni di interesse. In questo lavoro viene proposta una procedura analitica molto più pratica, che comporta l'analisi di una distribuzione normale bivariata a 5 parametri: medie e varianze delle due variabili e coefficiente di correlazione. La seguente relazione illustra come si possano ricavare il valore atteso e la varianza della distribuzione della coesione condizionale a $\Phi = \Phi_k$; la relazione è immediatamente applicabile in un foglio di calcolo.

Equazione 9: distribuzione condizionale di c dato Φk

$$E(C\,|\,\Phi=\varphi_k)=\mu_c+\rho\sigma_c\left(\frac{\varphi_k-\mu_\varphi}{\sigma_\varphi}\right)$$

$$VAR(C\,|\,\Phi=\varphi_k)=\sigma_c^2(1-\rho^2)$$

Dove:
ϕ_k = ϕ caratteristico (in questo caso, 5° percentile della distribuzione del campione)
μ_i è la media di i, dove i = ϕ, c
σ_i è la deviazione standard di i dove i = ϕ, c
ρ è il coefficiente di correlazione tra ϕ e c

Si pone in evidenza che, poiché i campioni di terreno prelevati e sottoposti a prove di laboratorio sono generalmente in numero ridotto o molto ridotto, è necessario o preferibile operare con l'ipotesi di varianza nota.

8. Esempio applicativo

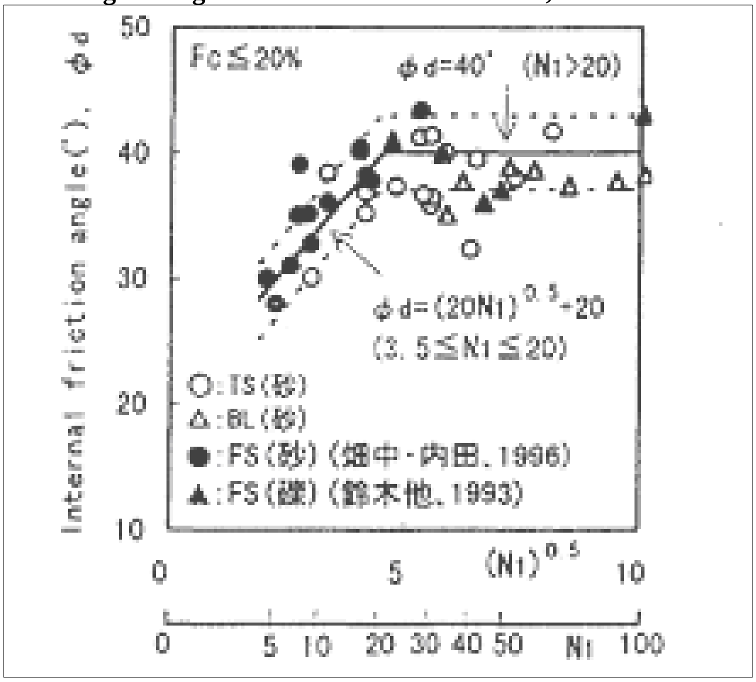
In un sito dove sono previste fondazioni superficiali (travi continue di 5 m di larghezza), sono state eseguite 3 prove SPT; la successione consiste in uno strato di sabbia da 0 a -6 m dal p.c., limi argillosi a profondità maggiore; la piezometrica è ubicata a -1 m dal p.c.. I dati relativi alle prove sono i seguenti:

Tabella 1

Identificativo prova	Profondità finale, m	NSPT	N'(60)	Φ °
N1	3.45	18	27	40
N2	3.45	19	28.5	40
N3	4.95	22	28.5	40

La legge di trasformazione utilizzata è quella di Hatanaka-Huchida [23] e Hatanaka et al. [24], che contiene un asintoto per N'(60) > 26. L'asintoto corrisponde al valore di Φ = 40, che in effetti è un valore medio compreso tra gli estremi di 37 e 43 circa, come visibile in figura 8 (dove N'≠ N'(60)). Per quanto attiene allo stato limite ultimo o verifica al collasso verticale della fondazione, la struttura di fondazione non è del tipo rigido e pertanto, mancando un segnale continuo di una prova penetrometrica, non siamo in grado di affermare se ci troviamo in condizioni di compensazione. Tuttavia, in considerazione della elevata larghezza della fondazione, possiamo con ragionevole certezza affermare che la potenziale superficie di rottura interesserà con elevata probabilità un volume di suolo grande rispetto alla scala di fluttuazione delle proprietà meccaniche del terreno. La media spaziale pertanto governa l'insorgere dello stato limite ultimo, e possiamo adottare come valore caratteristico il 5° percentile della media. Essendo in presenza di pochi dati, utilizziamo l'equazione 3; i tre valori di N'(60) sono molto simili, pertanto possiamo senza alcun dubbio adottare l'ipotesi di media del campione uguale a media della popolazione, con varianza nota pari al 10 % per litologie sabbiose. Il valore caratteristico risulta essere **Φ_k = 36.2°.** Si pone in evidenza che, nell'ipotesi di stato limite influenzato da valori locali (mancanza di compensazione spaziale), si dovrebbe applicare l'equazione 2, con 5° percentile calcolato sulla distribuzione del campione anziché quella della media; in tale caso, **Φ_k = 33.4°.** Il valore caratteristico così ottenuto verrà trasformato in valore di progetto secondo l'approccio utilizzato e inserito infine come input nell'equazione di verifica adottata per il calcolo della capacità portante:

Figura 8: grafico N'-Φ in Hatanaka et al., 1998



La verifica finale andrà effettuata trasformando Φ caratteristico in Φ di progetto applicando l'equazione 10, ed utilizzando il secondo nella formula scelta, in funzione anche dell'approccio di progetto utilizzato.

Equazione 10

$$\Phi_d = \tan^{-1}\left(\tan\frac{\Phi_k}{\gamma_m}\right)$$

Dove:

Φ_d è il Φ di progetto

γ_m è il fattore di sicurezza parziale

γ_m = 1.25 (approccio 1 combinazione 2)

γ_m = 1.00 (approccio 2)

9. Conclusioni

Il concetto dei valori caratteristici consiste nel caratterizzare un'intera distribuzione statistica dei parametri geotecnici di interesse mediante una stima puntuale a carattere cautelativo (modello semiprobabilistico). Tale stima puntuale viene generalmente individuata da un basso percentile, calcolato sulla distribuzione del parametro che influenza l'insorgere dello stato limite (ad esempio la media campionaria quando lo stato limite è influenzato dalla media spaziale dei dati o la media della distribuzione quando lo stato limite è influenzato dalla media locale dei dati). A tale scopo si adotta di solito il 5° percentile; in casi particolari, quali l'attrito negativo lungo il fusto di fondazioni profonde, la stima puntuale cautelativa può essere un alto percentile (il 95°). La cautela con la quale la particolare verifica geotecnica viene affrontata, rappresentabile dalla distanza del valore caratteristico dalla media dei dati, è in funzione del numero dei dati (che governa l'incertezza epistemica) della loro dispersione o variabilità. L'utilizzo della distribuzione della media dei dati in presenza di situazioni di compensazione (ampi volumi di terreno interessati dallo specifico stato limite) costituisce un meccanismo analogo a quello della riduzione della varianza utilizzato nella teoria dei campi aleatori. Il presente lavoro sintetizza i punti fondamentali nell'utilizzo dei valori caratteristici. Molte applicazioni di tipo operativo sono ancora da perfezionare; è stato comunque chiarito come il metodo statistico possa essere applicato in maniera rigorosa anche in presenza di pochi dati, contrariamente a quanto affermato da alcune voci critiche nei confronti del concetto dei valori caratteristici. Con dati molto scarsi (uno o due) è ancora possibile eseguire una analisi statistica rigorosa; i vincoli e le cautele da adottare rimangono i medesimi che esistevano nell'analisi tradizionale e prescindono dal modello semiprobabilistico. L'utilizzo del metodo statistico, una volta stabilite le regole fondamentali, è piuttosto semplice da attuare, con l'ausilio di comuni fogli di calcolo elettronici. Il metodo statistico va applicato in maniera ragionata e non può o deve sostituire il giudizio ed il discernimento tecnico. L'esperienza locale e regionale è un importante fattore che guida il processo di elaborazione statistica dei dati e che trova il suo fondamento teorico nell'analisi Bayesiana, non considerata nel presente lavoro. Permangono a tutt'oggi alcuni aspetti problematici nel concetto dei valori caratteristici, alcuni dei quali sono la modellazione dell'angolo di attrito con una distribuzione normale (o anche lognormale); il ruolo dell'angolo di attrito a volume costante e l'opportunità o meno di applicare allo stesso un fattore riduttivo γ_m [5]; l'individuazione dei valori caratteristici in presenza di due variabili correlate quali ϕ e c in prove triassiali o di taglio diretto. Alcuni di tali aspetti sono stati solo accennati in questo lavoro, altri saranno eventualmente oggetto di studi maggiormente approfonditi.

Riferimenti bibliografici

[1] Eurocode 7: Geotechnical design – Part 1: General rules, Final draft pr EN 1997-1, 2004, CEN

[2] Consiglio Superiore dei Lavori Pubblici, Istruzione per l'applicazione delle "Norme Tecniche per le Costruzioni" di cui al DM 14 gennaio 2008, bozza aggiornata al 7 marzo 2008, reperibile presso il link: http://www.rete.toscana.it/sett/pta/sismica/classificazione/index.htm

[3] Frank, R., Bauduin, C., Driscoll, R., Kavvadas, M., Krebs Ovesen, N., Orr, T., Schuppener, B.: Designers' guide to EN 1997-I, Eurocode 7: geotechnical design-general rules, 2004, Thomas Telford Ltd (series editor Gulvanessian)

[4] Bauduin, C.M., Determination of characteristic values. In: Smoltczyc (ed.), geotechnical engineering handbook, 2002, Wiley, volume I

[5] Bond, A., Harris, A., Decoding Eurocode 7, Taylor and Francis, 2008

[6] Fellin, W., Assessment of characteristic shear strength parameters of soil and its implication in geotechnical design. In: Fellin, W., Lessmann, H., Oberguggenberger, M., Vieider, R. (eds.), Analyzing uncertainty in civil engineering, 2005, Springer

[7] Kruse, B., Effects of the Determination of Characteristic Values of Soil Parameters, Robert Hack, Rafig Azzam, and Robert Charlier (Eds.): LNES 104, pp. 304–307, 2004 Springer-Verlag

[8] Lacasse, S., Nadim, F., Rahim, A., Guttormsen, T. R., Statistical description of the characteristic soil properties, Offshore Technology Conference, Houston, Texas, 2007

[9] http://areadocenti.eco.unicas.it/vistocco/download/materiale/tStudent.pdf

[10] http://www2.stat.unibo.it/roverato/index_file/Tavole_CLASED.pdf

[11] http://en.wikipedia.org/wiki/Student's_t-distribution

[12] Lacasse, S., Nadim, F., Uncertainties in characterizing soil properties, in: Uncertainty in the geologic environment: from theory to practice, Shackelford, C. D., Nelson P. P., Roth, M. J.S. (eds.), Madison, Wisconsin, 1996, ASCE special publication n° 58
[13] http://en.wikipedia.org/wiki/Stochastic_process
[14] http://en.wikipedia.org/wiki/Time_series
[15] http://it.wikipedia.org/wiki/Segnale_stocastico
[16] http://it.wikipedia.org/wiki/Processi_stocastici_ciclostazionari
[17] Schneider, H. R., Definition and determination of characteristic soil properties, Proceedings of the 14th International Conference on Soil Mechanics and Geotechnical Engineering , Hamburg 1997, Balkema, Rotterdam
[18] http://en.wikipedia.org/wiki/Linear_regression
[19] Duncan, M., Friction Angles for Sand, Gravel and Rockfill, Notes of a lecture presented at the Kenneth L. Lee Memorial Seminar Long Beach, California April 28, 2004
[20] Harr, M., Reliability-based design in civil engineering, Dover pbns 1996
[21] Cherubini, C., Reliability evaluation of shallow foundation bearing capacity on *c*' f' soils, Canadian Geotechnical Journal, 37, 264-269, 2000
[22] Van Alboom G., Mengè P., The derivation of characteristic values of shear strength parameters according to EC7, Geotechnical engineering for transportation infrastructure, Proceedings of the Twelth European Conference on Soil Mechanics and Geotechnical Engineering , Amsterdam- 1999, Balkema, Rotterdam
[23] Hatanaka M., Uchida A., Empirical correlation between penetration resistance and effective friction of sandy soils. Soils & Foundations, vol. 36 (4), 1-9, 1996, Japanese geotechnical society
[24] Hatanaka M., Uchida A., Kakurai M., Aoki M., A consideration on the relationship between SPT N-value and internal friction angle of sandy soils. J. Struct. Constr. Eng., AIJ, No. 506, 125-129, Apr., 1998

Appendice A – unbiased estimator per la deviazione standard: da

A1: **http://en.wikipedia.org/wiki/Standard_deviation**

Estimating population standard deviation from sample standard deviation

In the real world, finding the standard deviation of an entire population is unrealistic except in certain cases, such as standardized testing, where every member of a population is sampled. In most cases, the standard deviation is estimated by examining a random sample taken from the population. Using the definition given above for a data set and applying it to a small or moderately-sized sample results in an estimate that tends to be too low. The most common measure used is an adjusted version, the *sample standard deviation*, which is defined by

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}\,,$$

where $\{x_1, x_2, \ldots, x_N\}$ is the sample and $\overline{x}$ is the mean of the sample. The denominator $N - 1$ is the number of degrees of freedom in the vector $(x_1 - \overline{x}, \ldots, x_N - \overline{x})$.

The reason for this definition is that s^2 is an unbiased estimator for the variance σ^2 of the underlying population, if that variance exists and the sample values are drawn independently with replacement. However, s is *not* an unbiased estimator for the standard deviation σ; it tends to underestimate the population standard deviation. Although an unbiased estimator for σ is known when the random variable is normally distributed, the formula is complicated and amounts to a minor correction: see Unbiased estimation of standard deviation. Moreover, unbiasedness, in this sense of the word, is not always desirable; see bias of an estimator.

Another estimator sometimes used is the similar expression

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}\,.$$

This form has a uniformly smaller mean squared error than does the unbiased estimator, and is the maximum-likelihood estimate when the population is normally distributed.

A2 : David Vose, Risk analysis, a quantitative guide, Wiley 2000

La distribuzione dell’incertezza della media reale si calcola utilizzando al distribuzione di student:

$$\boxed{\mu = t_{n-1}^{0.95}\left(\frac{\hat{\sigma}}{\sqrt{n}}\right) + \overline{x}}$$

Dove $\hat{\sigma}$ è lo stimatore non distorto della deviazione standard reale, dato da:

$$\boxed{\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = s\sqrt{\frac{n}{n-1}}}$$

Attraverso sostituzione nell'equazione precedente si perviene dunque alla seguente:

$$\boxed{\mu = t_{n-1}^{0.95}\left(\frac{s}{\sqrt{n-1}}\right) + \bar{x}}$$

Appendice B – intervalli e livelli di confidenza per regressione lineare: da http://en.wikipedia.org/wiki/Linear_regression

Univariate linear case

We consider here the case of the simplest regression model, $Y = \alpha + \beta X + \varepsilon$. In order to estimate α and β, we have a sample $(y_i, x_i),\ i = 1, \ldots, n$ of observations which are, here, not seen as random variables and denoted by lower case letters. As stated in the introduction, however, we might want to interpret the sample in terms of random variables in some other contexts than least squares estimation.

The idea of least squares estimation is to minimize the following *unknown* quantity, the sum of squared errors:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

Taking the derivative of the preceding expression with respect to α and β yields the *normal equations*:

$$\begin{array}{l} n\ \alpha + \sum_{i=1}^{n} x_i\ \beta = \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i\ \alpha + \sum_{i=1}^{n} x_i^2\ \beta = \sum_{i=1}^{n} x_i y_i \end{array}$$

This is a linear system of equations which can be solved using Cramer's rule:

$$\hat{\beta} = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \bar{y} - \bar{x}\hat{\beta}$$

$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} y_i^2 - \frac{n(\sum_{i=1}^{n} x_i y_i)^2 + (\sum_{i=1}^{n} y_i)^2 \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{\sigma}^2 = \frac{S}{n-2}.$$

The covariance matrix is

$$\frac{1}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}$$

The mean response confidence interval is given by

$$y_d = (\alpha + \hat{\beta} x_d) \pm t_{\frac{\alpha}{2}, n-2} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_d - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

The predicted response confidence interval is given by

$$y_d = (\alpha + \hat{\beta} x_d) \pm t_{\frac{\alpha}{2}, n-2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_d - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$