

D. MAÏTI

M. C. TOUBOUL

Comparaison entre données réelles et données simulées en physique corpusculaire de haute énergie

Les cahiers de l'analyse des données, tome 4, n° 1 (1979),
p. 107-112

http://www.numdam.org/item?id=CAD_1979__4_1_107_0

© Les cahiers de l'analyse des données, Dunod, 1979, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

COMPARAISON ENTRE DONNÉES RÉELLES ET DONNÉES SIMULÉES EN PHYSIQUE CORPUSCULAIRE DE HAUTE ÉNERGIE [PHYS. COR. DEUTON]

par D. Maïti ⁽¹⁾ et M. C. Touboul ⁽²⁾

1 Comparaison variable par variable, et comparaison multidimensionnelle

Avant de préciser (§ 2) l'expérience de physique corpusculaire qui a fourni la matière de notre étude, décrivons en termes généraux la méthode de comparaison utilisée par nous.

Soit I_r un ensemble de faits individuels réels (notés i , ou i' , i'' ...): compte-tenu de l'application faite ici, il sera commode d'appeler I_r : ensemble des événements. Sur chaque événement i on a mesuré un ensemble V de variables v dites variables primaires et on note $v(i)$ la valeur de la variable v pour l'événement i . A côté des événements réels, un modèle théorique (ou semi-théorique: c'est-à-dire conjuguant des lois générales et des distributions empiriques) permet d'engendrer un ensemble I_s d'événements simulés (notés également i , ou i' ou i^1 etc.) dont chacun s'identifie à la suite (indiquée par v dans V) des valeurs $v(i)$ calculées pour lui.

Une première méthode pour confronter I_s à I_r , est de construire d'après l'un et l'autre de ces deux ensembles d'événements un histogramme de chacune des variables primaires v : c'est la comparaison variable par variable. Cependant un modèle théorique qui, de ce premier point de vue, apparaît satisfaisant peut ne pas résister à l'épreuve du calcul sur I_r et I_s d'un simple coefficient de corrélation entre deux variables primaires v^1 et v^2 ; *a fortiori* une analyse multidimensionnelle des deux tableaux de données - réelles $I_r \times V$ et simulées $I_s \times V$ - peut révéler deux structures incompatibles entre elles; de plus la statistique permet aussi d'engendrer par analyse des données primaires, des variables secondaires sur lesquelles apparaîtra manifestement l'inadéquation du modèle au réel (soit par la simple construction d'histogrammes, soit par une nouvelle application des méthodes multidimensionnelles). C'est d'un exemple de telle comparaison multidimensionnelle qu'il s'agira ici.

(1) Laboratoire de physique corpusculaire. Collège de France

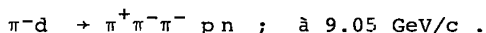
(2) Docteur ès sciences physiques L'NHE Pierre et Marie Curie (PARIS VI)

Le présent travail a déjà été présenté aux physiciens participant au : 3 Topical Meeting on multidimensional analysis of high energy data ; Nijmegen, 8-11 Mars 1978 ; cf Proceedings ; W. Kittel éd.

2 La réaction π -d à 9,05 GeV/c

Une série d'articles publiés dans ces *Cahiers* (cf [Phys.Cor.] Vol II, 1977, n°s 3, 4 ; Vol III, 1978, n° 1) a présenté au lecteur statisticien le type d'expérience considéré ici : en bref des corpuscules π^- constituant un faisceau de très faible densité mais de très haute énergie (équivalente, dans le système du laboratoire à près de 10 fois la masse d'un proton pour chaque π^-) pénètrent dans une chambre à bulle contenant du deutérium (hydrogène lourd) liquide, lequel sert à la fois de cible (les π^- peuvent réagir avec les noyaux notés d et appelés *deutons*) et de milieu d'observation (par formation de bulles matérialisant les trajectoires, qui sont photographiées ; puis mesurées pour en déterminer par le calcul, les variables cinématiques qui sont nos variables primaires ; cf [Phys.Cor.] § 2.3 ; Vol II, p. 315).

Les données analysées proviennent d'une expérience réalisée et analysée conjointement par deux laboratoires (LPNHE : Paris & INFN : Bari ; cf e.g. Baubillier & al. (*) Paris ; Armenisse & al. Bari : *Pion and Rho exchange model for the diffractive reaction $\pi^-d + \pi^-p^0d$ at 9 GeV/c* ; exposé à la Conférence internationale de Budapest ; 1977). On considère 4218 événements de la réaction :



Au physicien cette formule suggère immédiatement le commentaire suivant : le deuton d, état lié d'un proton p et d'un neutron n a fourni les deux nucléons (p et n) émergents séparés ; au lieu d'un méson π^- incident, on a trois mésons émergents (deux négatifs et un positif) , la

formation de ces mésons étant générale dans les réactions à ce niveau d'énergie. Plus précisément, la mécanique quantique décrit le deuton d, comme un état faiblement lié de deux nucléons, un p et un n, animés d'un mouvement rotatif, dit *mouvement de Fermi*, suivant la loi semi-empirique de Hulthen. D'où l'hypothèse : dans chaque événement, il y a eu choc et réaction entre le π^- incident et l'un des deux nucléons du d (soit le n soit le p) dit nucléon *acteur* ; tandis que l'autre nucléon dit nucléon *spectateur*, peu influencé par la réaction parce que faiblement lié à l'*acteur*, poursuit dans la gerbe émergente une trajectoire rectiligne dont la vitesse révèle le *mouvement de Fermi* (comme le fait une mesure , au sens qu'a ce terme en mécanique quantique, portant sur la fonction d'onde du deuton : laquelle n'est évidemment pas celle d'un mouvement rectiligne pur... ; mais une superposition d'ondes décrivant de tels mouvements).

Si l'on accepte cette hypothèse, le spectateur doit pouvoir être reconnu entre les deux nucléons émergents (n et p) comme celui qui a la plus faible vitesse ; la distinction étant particulièrement nette si l'un des nucléons a une vitesse comprise dans les limites que donne la formule de Hulthen ; tandis que l'autre s'en écarte grandement . De plus dans la mesure où un modèle fondé sur l'interférence entre deux termes principaux symbolisés par les graphes de Feynman de la figure 1, peut décrire la réaction entre un méson π^- et un nucléon (soit n soit p), ce modèle combiné avec la formule de Hulthen permettra de simuler les réactions π^-d :

L'étude physique avait en fait pour objet principal la réaction π^-n ; c'est pourquoi on a simulé quelques milliers d'événements de la réaction π^-d , avec *proton spectateur* (donc *neutron acteur* ; et réaction π^-n). Bien qu'en première analyse l'accord entre l'expérience réelle et le modèle ait paru satisfaisant, le physicien (M.C. T.), a eu recours au statisticien (D. M.) d'abord pour revoir la séparation des 4218 événements réels entre réactions sur neutron acteur et réactions sur proton spectateur ; et ensuite pour comparer données réelles et données simulées.

(*) dont l'un de nous : M.C. T.

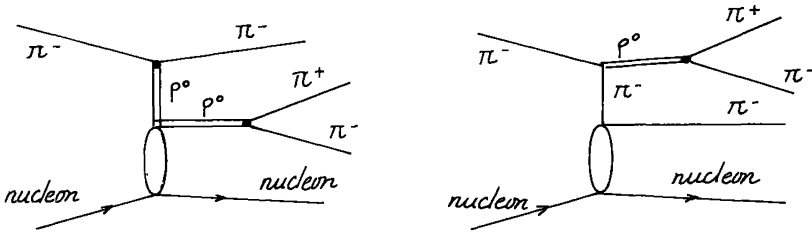


Figure 1 : Les deux graphes de Feynman utilisés dans le modèle de la réaction π -nucléon : on notera que ces deux graphes comportent l'existence intermédiaire d'un méson ρ^0 ; sur le rôle de ces graphes dans les modèles, cf [Phys. Cor.] § 3.5 ; Vol II p. 327.

3 Données primaires et données secondaires

Chaque événement, réel ou simulé, est décrit par les quadrimoments des cinq particules émergentes ; à partir de ces données on calcule d'une part, trois variables physiques primaires dont l'importance est manifeste pour notre problème ; et d'autre part trois variables secondaires introduites par le statisticien comme des caractéristiques globales de la gerbe émergente. Enfin vu leur hétérogénéité, ces 6 variables numériques sont codées sous forme logique (en 0,1) en divisant en 10 intervalles consécutifs l'intervalle de variation des 6 variables continues (codage classique en a. des correspondances : cf e.g. B. Ghermani et C. & M. Roux ; in *Cahiers* Vol II n° 1 pp. 115 sqq).

Voici donc le détail des 3 variables primaires :

PN : valeur absolue de la quantité de mouvement du neutron émergent mesurée dans le système du laboratoire : $PN = \|\vec{p}_n\|$;

PP : id. pour le proton émergent : $PP = \|\vec{p}_p\|$

RI : rapport PN/PP.

Quand aux variables secondaires, leur calcul repose sur la notion de transversalité propre exposée par P. Lutz et D. Maïti dans un article des ces cahiers (Vol III n° 4, pp 449-458). Bornons-nous à rappeler que toute gerbe de particules (ou toute gerbe partielle), considérée dans le système de son propre centre des masses est décrite par un nuage ou ensemble de vecteurs tridimensionnels (vecteurs usuels : les quantités de mouvement des particules) : elle peut donc être soumise à une analyse factorielle, i.e. à la recherche des axes principaux d'inertie et des moments $\lambda_1, \lambda_2, \lambda_3$. Si toutes les vitesses étaient colinéaires, on aurait un nuage réduit à une droite : $\lambda_1 \neq 0$; $\lambda_2 = \lambda_3 = 0$; mais en général, il y a une direction principale (1° axe) qui définit la tendance longitudinale globale de la gerbe, avec une dispersion transversale que l'on peut mesurer par le quotient $(\lambda_2 + \lambda_3)/\lambda_1$; ce quotient, ou *transversalité propre* est généralement faible dans les réactions à très haute énergie. Ceci étant rappelé, disons qu'on a calculé :

RN : transversalité $((\lambda_2 + \lambda_3)/\lambda_1)$ pour le système des trois pions et du neutron

RP : transversalité pour le système des trois pions émergents et du proton

R2 : rapport RN/RP.

Pour justifier le choix de ces mesures, il faut considérer l'hypothèse physique initiale : si le neutron est acteur PN est élevé (car le neutron a reçu le choc du π^- incident), PP est faible (réduit au mouvement de Fermi des nucléons constituant le deuton cible) et le rapport R1 est donc élevé. A cette hypothèse, l'analyse des réactions à haute énergie (toujours fortement longitudinales, comme on vient de le rappeler) suggère d'en adjoindre une autre : si le neutron est acteur, c'est l'événement $\pi^- n \rightarrow \pi^- \pi^- \pi^+ n$ qui correspond à une réaction réelle donc RN est faible ; tandis que le système $\pi^- \pi^- \pi^+ p$ formé en adjoignant aux trois mésons π le proton qui n'est pas intervenu dans la réaction est une gerbe fictive dont la transversalité propre n'a pas de raison d'être faible : donc RP est élevé, et le rapport R2 = RN/RP est faible. Pour une réaction sur proton acteur avec neutron spectateur, c'est le contraire : R1 faible, R2 élevé. Nous concluons donc que selon les deux hypothèses considérées les rapports R1 et R2 devraient varier en sens opposé.

4 Analyse des données réelles

En vue du codage logique, l'intervalle de variation de chacune des six variables {PN, PP, R1, RN, RP, R2} est divisé en dix intervalles consécutifs - notés respectivement PN1, PN2, ..., PN10 ; PP1, PP2, ..., PP10; ...; R21, R22, ..., R210 - définissant les modalités de la variable correspondante : e.g. PN1 signifie PN le plus faible, ..., PN10 signifie PN le plus fort ; les bornes des intervalles sont choisies pour que dans chacun d'eux rentre un dixième (i.e. 412 ou 413) des 4218 événements réels.

A partir de l'ensemble J des 60 (6 fois 10) modalités ainsi définies, on construit comme il est d'usage (cf Ghermani et coll. ; *loc. cit*), le tableau de cooccurrence, ou tableau de Burt 60 x 60 : avec $k(j, j')$ = nombre des événements rentrant à la fois dans les modalités j et j' .

Ce tableau a été soumis à l'analyse factorielle, et nous commenterons brièvement le graphique du plan 1 x 2. Mais, à la vérité, le tableau lui-même - plus précisément le sous-tableau 10 x 10 croisant les modalités des rapports R1 et R2 - suffit à montrer combien les données réelles s'accordent peu avec ce que nous attendions (cf tableau 1). Les rapports R1 et R2 ne varient pas dans des directions opposées ; bien au

	R21	R22	R23	R24	R25	R26	R27	R28	R29	R210
R11	118	146	81	41	23	8	2	3	0	0
R12	52	84	86	94	50	34	14	5	1	1
R13	30	49	80	109	71	48	18	13	3	0
R14	32	38	61	62	84	71	36	23	12	3
R15	23	34	41	33	67	72	66	49	29	7
R16	19	14	22	25	31	48	77	88	58	40
R17	23	14	20	15	24	45	58	79	97	47
R18	14	19	16	14	19	36	59	68	88	89
R19	36	13	10	11	15	11	37	44	77	167
R110	74	10	4	17	37	48	54	49	56	175

Tableau 1 : Bloc R1 x R2 du tableau de Burt construit pour 4218 événements réels

	R21	R22	R23	R24	R25	R26	R27	R28	R29	R210
R11										
R12										
R13										
R14										
R15				1				1		
R16				0				0	2	
R17	5			0			1	0	4	1
R18	10	3	2	1			6	2	3	2
R19	130	28	2	15	8	16	18	41	39	48
R110	771	94	38	45	46	65	81	99	117	238

Tableau 2 : Bloc R1×R2 du tableau de Burt construit pour 1999 événements simulés, en conservant les bornes de classes déterminées pour des événements réels.

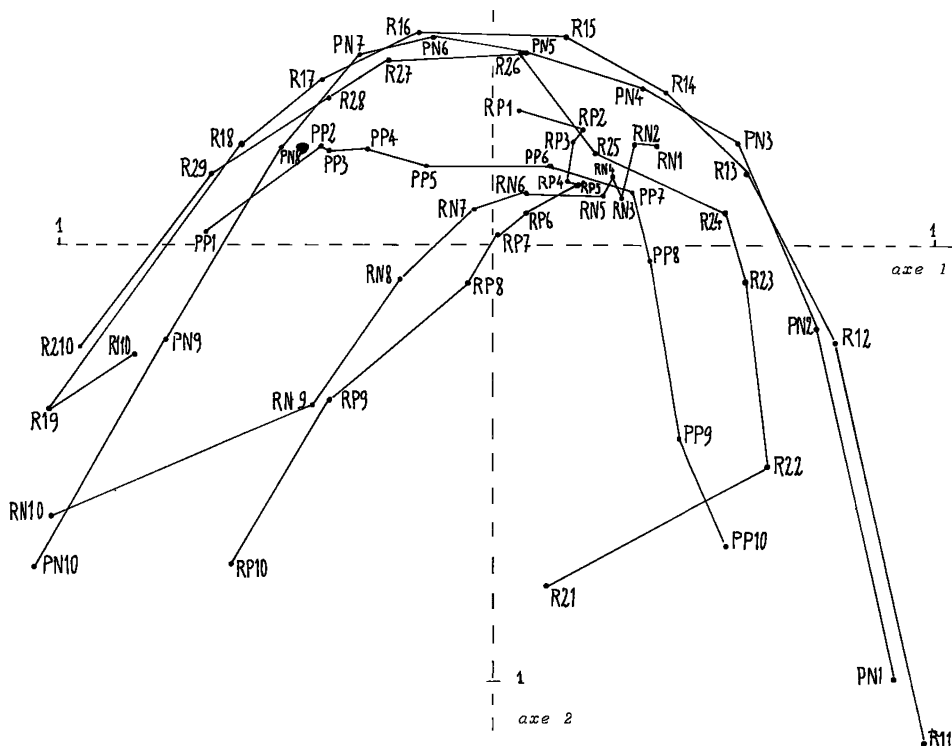


Figure 2 : Plan 1 × 2 issu de l'analyse du tableau de Burt 60 × 60 calculé d'après 4218 événements réels.

contraire, la plupart des événements tombent au voisinage de la diagonale. On trouve seulement dans le coin inférieur gauche quelques événements (dans deux cases $74 + 36 = 110$; et peut-être quelques autres...) pour lesquels une valeur élevée de R1 est associée à une faible valeur de R2: combinaison de valeurs que nous pensions *a priori* devoir caractériser les événements pour lesquels le neutron est acteur. Le coin opposé (coin supérieur droit : R1 faible ; R2 élevé) qui devrait correspondre aux événements sur proton acteur, est à peu près vide.

Selon le Professeur L. Van. Hove (cf discussions au colloque de Ni-mègue) les résultats recensés dans le tableau 1, s'accordent avec le modèle physique suivant, seul justifié selon lui :

... La réaction π^-d est fortement dominée par la dissociation diffractive du π . Bien que le deuton d soit dissocié, il ne convient pas de distinguer entre nucléon spectateur et nucléon acteur : d se dissocie parce que son énergie de liaison est très faible vis-à-vis de l'énergie de la réaction.

Quant à l'analyse factorielle, nous dirons que le plan 1×2 offre une vue synthétique de résultats déjà apparents à la lecture du tableau de Burt 60×60 . La plupart des variables ont avec PN et R1 une forte corrélation positive : en ce sens que leurs dix modalités dominent une courbe qui suit à peu près celle de PN ou de R1 ; (avec simultanément dans le tableau de Burt des blocs 10×10 où prédominent les cases voisines de la diagonale principale) ; seul PP varie dans la direction opposée. Il est particulièrement remarquable que RN et RP (transversalité propre des sous-gerbes $n3\pi$ et $p3\pi$ respectivement) diffèrent peu l'un de l'autre. La position remarquable du point R21, qui s'écarte de R22 dans la direction de R110, (se plaçant ainsi à l'intérieur du croissant parabolique dessiné par les autres modalités), correspond aux quelques 110 événements du coin inférieur gauche du tableau 1 ; événement associant R21 à R110 (ou R19).

5 Comparaison des données simulées aux données réelles

Il suffit de comparer le tableau 2 construit à partir des données simulées en conservant les bornes de classes déterminées sur les 4218 événements réels recensés dans le tableau 1. Que de différences entre les deux tableaux ! D'abord, dans le tableau 2, les lignes correspondant aux faibles valeurs de R1 (R1 = PN/PP) sont vides : ce qui est normal, puisque la simulation concerne seulement des événements engendrés suivant le modèle du *neutron acteur* (donc avec un PN plus élevé que PP : le nucléon acteur émerge avec une vitesse supérieure à celle du spectateur). Mais ceci dit, on devrait avoir dans le tableau 2, une partie inférieure peu différente de celle du tableau 1 : il n'en est rien. Notamment le coin inférieur gauche du tableau 2 contient environ la moitié des événements simulés ; tandis que dans le tableau 1 on a trouvé 100 événements exceptionnels seulement : soit moins de 5% des 2100 événements réels recensés dans la moitié inférieure du tableau 1. Ainsi les données simulées s'accordent bien mieux que les données réelles avec ce que nous attendions : association R1 fort - R2 faible, pour de nombreux événements interprétés comme une réaction entre π^- incident et neutron acteur. En effet la simulation a été faite suivant l'hypothèse d'un seul nucléon acteur, hypothèse qui selon L. Van Hove, n'est pas acceptable ; et que l'analyse des données réelles nous apparaît rejeter.

Nous concluons donc que grâce principalement au calcul des variables secondaires introduites au § 3, l'analyse statistique a nettement écarté une hypothèse physique initiale, et avec elle, le modèle de simulation que celle-ci suggérait.