

La simulation numérique 1/2

Introduction

Cela a déjà été expliqué de manière détaillée dans différents podcasts, pour comprendre le monde qui nous entoure et arriver à savoir ce qui nous a amené où nous en sommes et arriver à prévoir ce qu'il va arriver, il est nécessaire de le modéliser.

J'ai trouvé un certain nombre de définitions de ce que peut-être un modèle, et elles diffèrent en fonction des domaines d'étude (mathématique, économie, physique, etc). Ce qui en ressort, de mon point de vue, est que le but du modèle est de représenter la réalité en éliminant les détails difficiles à reproduire (ou n'étant pas important pour la simulation ou la résolution du problème) et en se concentrant sur les seuls éléments jugés importants pour l'étude.

Dans l'épisode sur les objets mathématiques, Robin l'explique très bien quand il parle du pont de Koënigsberg. L'important est uniquement le graphe (le modèle) et pas la réalité (les ponts, les rues, etc) car il réduit la réalité aux seuls éléments importants pour l'étude que l'on souhaite faire.

Le modèle ne constitue pas la réalité, il la représente seulement. Comme l'énonce l'adage d'Alfred Korzybski : "La carte n'est pas la territoire" ¹.

C'est pour cela que les modèles n'ont pas forcément à être le plus précis possible ou à intégrer tous les phénomènes possibles et imaginables. Pour étudier le comportement d'un fluide (en l'occurrence l'air) autour d'une aile d'avion, on ne s'intéresse souvent qu'à l'écoulement de l'air autour du profil de l'aile. Ainsi le modèle va être une coupe transversale de l'aile autour de laquelle l'air va s'écouler. On va ensuite observer le comportement de l'écoulement autour de ce profil d'aile en fonction de différents paramètres de vitesse, d'altitude, de forme de profil, etc.

Le modèle dans ce cas-là va être constitué du profil de l'aile et des lois qui vont régir le fluide (l'air ici) qui va s'écouler autour de cette aile.

Il est important de comprendre que les modèles ont leur domaine de validité, de la même manière que la relativité générale et la mécanique quantique ont leur domaine d'application respectif au-delà desquelles elles ne s'appliquent plus.

Les modèles permettent ainsi l'étude de la réalité dans un cadre défini dans lequel ils la représenteront bien et dans lequel nous aurons des données qui les valideront.

Grâce à cette validation, les modèles seront réputés fiables, ils permettront de reproduire des phénomènes réels, dans ce cadre défini je le répète, et ils pourront être utilisés pour la simulation et la prédiction.

¹http://fr.wikipedia.org/wiki/Alfred_Korzybski

Je sais que part le passé, il y a eu tout un échange sur les modèles, et notamment leur limites (un HS d’ailleurs), est ce que vous pourriez donner votre point de vue sur leur usage ?

L’intérêt des modèles ² est donc, dès lors qu’ils sont corrects, de pouvoir permettre différentes choses :

- La simulation des phénomènes afin de prévoir comment ils vont se comporter dans le futur ou soumis à des contraintes spécifiques
- La simulation d’interaction entre différents phénomènes où il est nécessaire de savoir comment ceux-ci interagissent
- Valider que les modèles représentent bien ce qu’ils sont censés modéliser

Comme l’explique Wikipedia ^{3 4} : ce que l’on appelle simulation correspond à l’ensemble constitué par un modèle, les paramètres et contraintes, et les résultats obtenus.

Les maquettes, prototypes, etc. peuvent ainsi être considérés comme des modèles analogiques et les essais, tests, manœuvres, etc. comme des simulations analogiques.

Et pendant très longtemps il s’agit du seul type de simulation qui était à disposition des scientifiques. Ces simulations ont permis de reproduire et de comprendre un grand nombre de phénomènes en laboratoire.

Les équations de leur côté sont plutôt des simulations numériques et aujourd’hui ce terme s’applique essentiellement aux modèles et simulations réalisés sur ordinateur et basés sur ces équations.

La simulation numérique est apparue grâce à la conjonction de deux facteurs :

- Dans certains cas, la simulation “analogique” était :
 - compliquée comme l’étude de problèmes multi-physiques en aéronautique avec de l’eau de la glace de l’air, des profils d’ailes spécifiques, les différentes techniques de dégivrage et l’impact de tout ça sur le vol,
 - chère parce que les objets d’études étaient onéreux à produire comme une voiture donc on souhaite comprendre la résistance aux crashes ou nécessitaient des matériels ou équipement onéreux comme le LHC,
 - longue à réaliser : les durées nécessaires pour observer les phénomènes étudiés étaient longues comme certaines réactions chimiques par exemple ou les préparations des expériences tellement compliquées qu’elles prenaient le pas sur l’expérience elle-même comme dans les cas où le matériel à tester est si rare qu’il faut s’assurer que l’expérience sera sans faille.

²http://fr.wikipedia.org/wiki/Mod%C3%A8le_math%C3%A9matique

³http://fr.wikipedia.org/wiki/Simulation_de_ph%C3%A9nom%C3%A8nes

⁴http://fr.wikipedia.org/wiki/Mod%C3%A8le_math%C3%A9matique

- voire impossible (allez simuler en laboratoire l’univers ...)
- Les ordinateurs qui sont apparus dans les années 50 permettaient une reproduction d’“expérience” de manière automatique, rapide et exact (en accord avec le modèle utilisé)

Il faut bien noter que dans la simulation numérique nous étudions autre chose que le réel. Il faut être clair sur le sujet. Quand on souhaite utiliser un ordinateur pour simuler, il est indispensable de pouvoir définir un modèle théorique qui soit informatisé. Et comme on ne peut représenter le réel grâce à un ordinateur qui ne peut travailler que sur des nombres finis d’éléments, on va donc travailler avec des approximations du réel que l’on essaiera de rendre les plus fidèles possible. On va ainsi discrétiser une droite en prenant des points, on aura des nombres qui ont des tailles finies en mémoire, etc et tout cela amènera de l’imprécision.

Juste pour la précision, discrétiser signifie passer du continu (une ligne contient un nombre infini de point) au discret (un nombre fini de point suffit à décrire la ligne pour avoir quelque chose de représentatif). Quand il s’agit de simulation numérique, la discrétisation d’un milieu continu est un maillage, c’est à dire la représentation géométrique d’un domaine par des éléments finis et définis.

Mais si on regarde bien, de la même manière on ne mesure le réel que de manière approximative (avec une certaine précision que l’on jugera satisfaisante pour l’étude que l’on mène).

Un gros travail va donc être indispensable de réaliser quand on mettra en place des simulations numériques: celui de s’assurer que le modèle que l’on utilise est assez fidèle à la réalité, pour ce que l’on souhaite étudier, et que la précision que l’on recherche est suffisante par rapport à celle que l’on a lorsqu’on étudie le réel.

Pour peu que l’on soit conscient des limites des simulations, qu’on les utilise de manière “raisonnée” et dans les bons contextes avec toujours un œil critique sur les résultats obtenus, elles pourront être un très bon outil pour les chercheurs et les ingénieurs dans leurs recherches.

Un bon exemple d’un domaine où l’expérimentation et les tests coûtent chers et où l’on a cherché à mettre de la simulation numérique est celui de l’industrie de l’aéronautique. Pour faire des simulations “analogiques”, le domaine dispose grosso modo de deux grands moyens :

- On peut faire des tests en soufflerie
- On peut faire des tests en conditions réelles

Expériences et tests en soufflerie

Il s’agit ici de tester l’écoulement de fluides autour d’éléments dont on veut comprendre le comportement. On se sert la plupart du temps de fumée pour

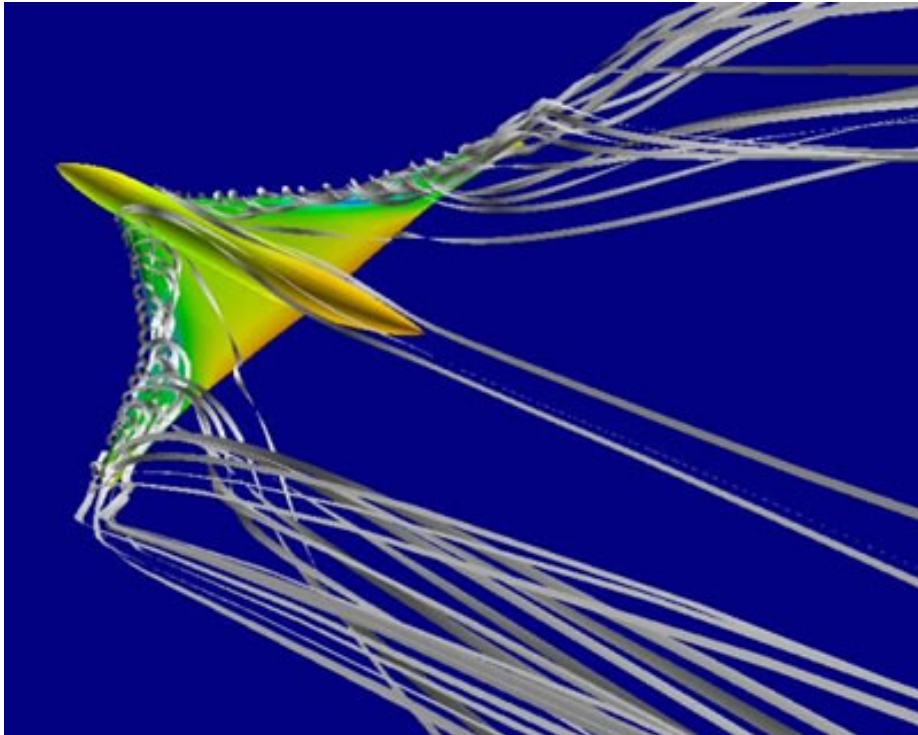


Figure 1: Simulation numérique de l'aérodynamique de l'hypersustentation de l'avion de transport supersonique Epistle (DAAP - Gérard Carrier) <http://hathor.onera.fr/images-science/simulation-numerique/sst-epistle.php>

“voir” les écoulements. Les éléments testés peuvent être instrumentés (les objets sont équipés de capteurs qui vont permettre de connaître différentes grandeurs physiques comme la température, la pression, etc) pour comprendre les températures ou les contraintes auxquels ils sont soumis.

La réalisation de tests en soufflerie nécessite de disposer de soufflerie suffisamment grande pour arriver à tester les systèmes en grandeur nature : voiture, éléments d’avion. Mais aussi de faire un très grand nombre d’essai pour arriver à tester les différentes conditions que l’on souhaite observer. Il est même compliqué voire impossible de tester des avions en soufflerie : soit ils sont trop volumineux et alors on ne teste que des composants (les ailes par exemple), soit cela se fait avec des réductions.

Seules certaines composantes des écoulements vont pouvoir être testé, on ne peut en effet qu’envoyer des filets de fumée afin de pouvoir voir comment ils se comportent.

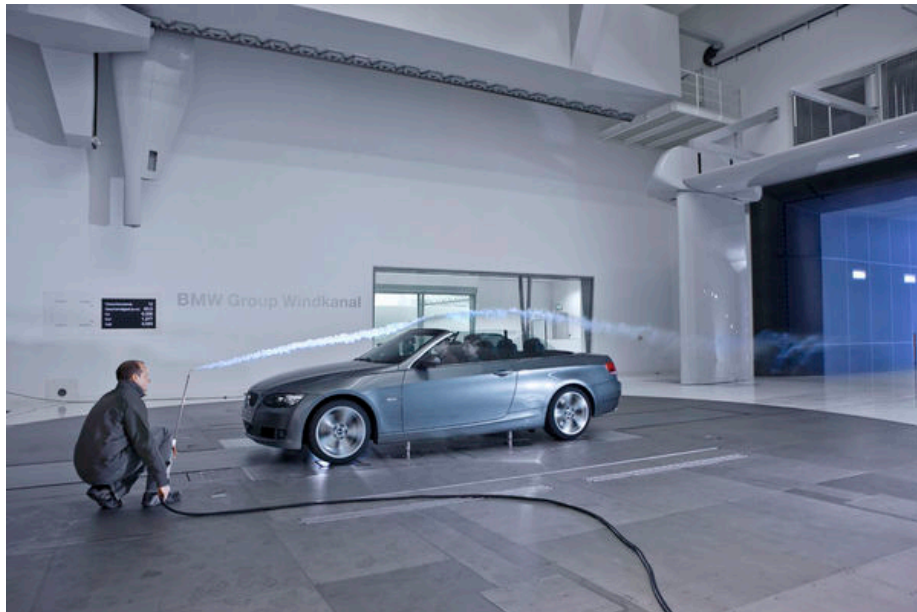


Figure 2: Vu de face d’un BMW en soufflerie <http://www.usinenouvelle.com/article/bmw-joue-a-fond-la-carte-ecolo.N70485>

On voit donc que cela peut devenir compliqué de tester des conditions extrêmes de vol comme par exemple (températures basses, écoulement d’air, gouttelettes d’eau pour simuler des nuages, etc).

En grandeur réelle

Dans ce cas-là, c'est encore plus complexe. Déjà il est nécessaire de construire des avions en taille réelle pour qu'ils puissent être pilotés et cela coûte extrêmement cher! Je ne prétends pas que cela ne soit pas fait, mais dans une mesure qui ne permet pas de pouvoir s'en servir comme source principale d'information concernant le comportement des avions construits. En effet, et si le design possédait une faille majeur qui, une fois en vol provoquerait un crash ?

Il apparaît donc indispensable de pouvoir simuler sur ordinateur des avions pour comprendre comment cela se passe. Une fois que tout est validé par la simulation numérique, il peut ensuite être temps de faire des tests de vérification d'éléments de fuselage et enfin des tests en vol pour valider que ce qui était attendu pour l'avion au global. Et là, je ne parle même pas des tests de crash d'avion comme cela pourrait être le cas avec voiture comme je l'ai expliqué plus haut.

Mais l'aéronautique reste un domaine où l'expérimentation, bien que très chère reste possible. Il y a des domaines où faire des simulations analogiques est tout bonnement impossible.

Domaines où l'expérience est impossible

La météorologie

Prenons un autre domaine où la simulation est indispensable : la météorologie ⁵ ⁶. Ici la reproductibilité n'est même plus possible. L'idée est plutôt de pouvoir prévoir le temps. Réaliser des tests à échelle réduite ne me semble pas possible, ni même pertinent. En effet il est indispensable de pouvoir simuler l'atmosphère, son interaction avec les sols, avec l'océan, l'impact de la nuit et du jour, etc, etc. Quand on parle de météorologie, certaines personnes tiquent un peu quand on dit que c'est une science. Il est vrai que l'on observe souvent nos chers présentateurs météo annoncer de la pluie en Bretagne à longueur de temps et on entend régulièrement certains quimperois expliquer à corps et à cris que non, "il faisait grand beau ce jour là, et qu'ils en ont ras la crêpe". En fait cela vient d'au moins deux phénomènes :

- **La discrétisation et le maillage choisi.** Pour réaliser des simulations sur un ordinateur (un système digital et pas analogique qui fonctionne donc par étape discrète), il faut discrétiser son domaine d'étude (le découper en morceaux) et produire un maillage (une grille grosso modo) dont les intersections seront les points auxquels nous chercherons les valeurs de la température (pour faire simple). Sauf que vouloir simuler tout un pays,

⁵http://fr.wikipedia.org/wiki/Pr%C3%A9vision_m%C3%A9t%C3%A9orologique#Pr.C3.A9vision_num.C3.A9rique

⁶http://fr.wikipedia.org/wiki/Pr%C3%A9vision_num%C3%A9rique_du_temps

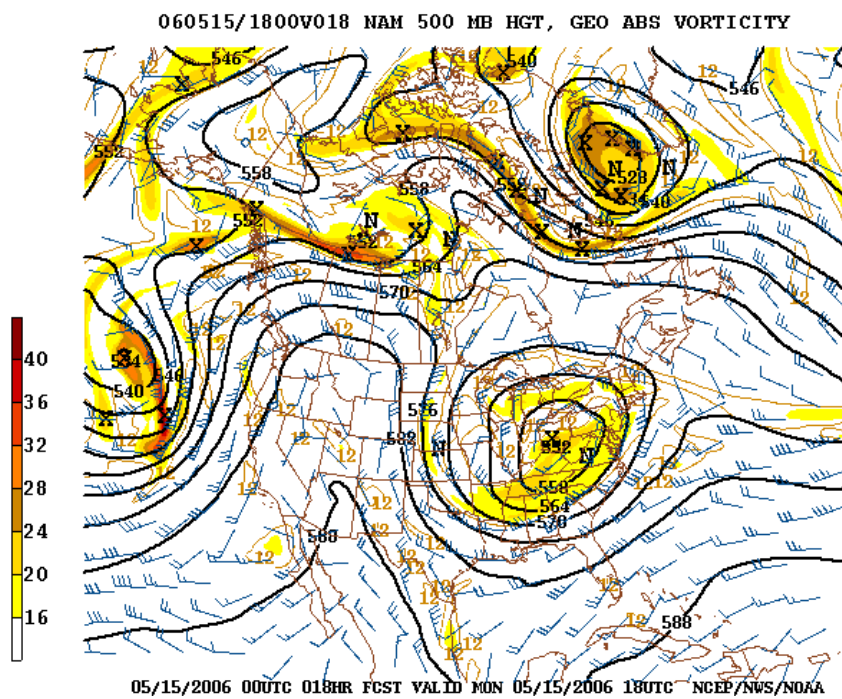


Figure 3: Pr vision m t orologique r alis e par mod le de pr vision am ricain NAM http://commons.wikimedia.org/wiki/File:NAM_500_MB.PNG

voire plus pour prendre en compte la mer ou des choses comme les montagnes, etc. donc de grands territoires, c'est compliqué. Afin de pouvoir aller vite, la distance entre deux points du maillage peut-être de plusieurs kilomètres. Le problème c'est que des "micro-climats" ou des orages par exemple vont avoir de forts impacts à de plus grandes échelles. Pour résoudre cela on peut "paramétriser" les mailles, c'est-à-dire ajouter des bouts de modèles à ces échelles qui vont prendre en compte certains phénomènes. Cela reste néanmoins approximatif et pas forcément très efficace.

Je ne sais pas si tout le monde a suivi jusqu'ici, est ce qu'il y a des questions sur ces problèmes? Je peux vous donner quelques exemples de maillage pour que vous puissiez voir ce que cela peut donner pour des modèles utilisés "dans la vraie vie".

Pour donner un ordre d'idée de maillages utilisés, voici ce que Wikipédia donne comme info ⁷ : "L'IFS (Integrated Forecasting System), calculé par le Centre européen de prévision météorologique à moyen terme à Reading en Grande-Bretagne, donne une prédiction pour l'Europe avec une maille mondiale de 25 km de côté. Météo-France de son côté utilise actuellement pas moins de trois modèles numériques, telles trois boîtes imbriquées les unes dans les autres pour émettre ses bulletins. Arpège (Action de recherche petite échelle /grande échelle) utilise le résultat de l'IFS et refait le calcul avec une maille de 25 km sur la France pour une échéance de un à trois jours, mais moins précise ailleurs. Aladin (Aire limitée et adaptation dynamique) reprend les résultats pour l'Europe de l'Ouest avec une précision de 10 km. Depuis 2008, le modèle AROME⁸ (Application de la recherche à l'opérationnel à mésoéchelle) calcule sur une maille de 2,5 km (cent fois plus précise en surface que le modèle Arpège et près de seize fois plus qu'Aladin). Le calcul intègre en continu et réajuste les prévisions à partir des informations des stations météorologiques, navires, bouées, avions, radar, satellites... et intègre aux données de base : les vents, les précipitations et l'humidité de l'air fournies par les satellites GPS. Le modèle évalue aussi la fiabilité de la prédiction". Un petit récap sur les trois modèles peut-être trouvé ici ⁹

- **La couverture en données du domaine considéré et l'erreur sur les mesures de ces données.** Il apparaît clair, que les zones géographiques terrestres sont plus couvertes et permettent d'avoir plus de données, donc une plus grande précision sur les conditions initiales du

⁷http://fr.wikipedia.org/wiki/Pr%C3%A9vision_m%C3%A9t%C3%A9orologique#Pr.C3.A9vision_num.C3.A9rique

⁸<http://www.cnrm.meteo.fr/IMG/pdf/arome2007.pdf>

⁹http://www.mementodumaire.net/wp-content/uploads/2012/07/Meteo_magazine.pdf

domaine. Par contre, pour les océans c'est plus compliqué. Et même avec les informations que les balises en mer peuvent fournir (sur la température ou les courants marins), on se retrouve avec beaucoup moins de données. Pour palier cela, on va adjoindre à ces observations des résultats précédant pour essayer de se rapprocher au mieux de la réalité. Un point important à savoir quand on fait de la simulation temporelle comme pour les prévisions météo, c'est que l'on fait évoluer le système à partir de ses conditions initiales et donc plus celles-ci sont précises, plus le résultat final sera précis. Et inversement. Ajouté à cela les erreurs numériques des modèles (des approximations de la réalité) et celles des ordinateurs, on se retrouve avec un système qui peut totalement diverger et ceci vers des solutions qui pour des simulations indépendantes avec des écarts minimes au niveau des données d'entrée peuvent être totalement différentes ! Cet aspect de "chaos" (faudrait faire un dossier sur le sujet, c'est juste un truc de fou) a notamment été mis en exergue par Poincaré pour la simulation à N corps du système solaire en 1890 (Il y a une jolie histoire sur ce qui a amené Poincaré à découvrir cela :)) et ensuite par Edward Lorenz¹⁰ pour la météo en 1963. C'est d'ailleurs lui qui est à l'origine de la fameuse métaphore du papillon et les simplifications qu'il avait du faire pour tenter des simulations plus simple de météo ont abouti à la découverte des "attracteurs étranges" qui font qu'en fait, pour presque toutes les conditions initiales possibles, on se retrouve attiré dans un schéma qui sera toujours le même, sans pour autant savoir, pour une étape donnée, à quelle endroit du schéma on se trouvera^{11 12}. Une jolie façon de le dire est la manière suivante¹³ :

D'où qu'on parte, on se retrouve toujours sur l'attracteur, c'est le côté prévisible de l'évolution. Où se retrouve-t-on exactement sur l'attracteur? Il est impossible de répondre à la question, c'est le côté imprévisible de l'évolution

En 1890, Poincaré publie le fameux mémoire intitulé : "Sur le problème des trois corps et les équations de la dynamique", qui lui vaudra le prix du roi Oscar, roi de Norvège et de Suède et passionné de mathématiques. L'histoire est célèbre : le mémoire lauréat comportait une erreur détectée par le jeune mathématicien Phragmén alors qu'il prépare le manuscrit pour l'imprimeur. Cette erreur obligera Poincaré à procéder à de profonds remaniements dans son mémoire, et aussi à rembourser les frais d'impression du premier mémoire, une somme supérieure de quelque mille couronnes au prix qu'il avait reçu. Mais cette erreur fut féconde, car en lieu et place de la stabilité du

¹⁰http://fr.wikipedia.org/wiki/Edward_Norton_Lorenz

¹¹http://fr.wikipedia.org/wiki/Syst%C3%A8me_dynamique_de_Lorenz

¹²http://fr.wikipedia.org/wiki/Th%C3%A9orie_du_chaos

¹³http://math.cmaisonneuve.qc.ca/alevesque/chaos_fract/Attracteurs/Attracteurs.html

Système solaire, Poincaré découvrit le chaos potentiel caché dans les équations de la dynamique.

Il s'agit ici d'un des domaines dans lequel la théorie du Chaos peut dire des choses, je ne sais pas si vous avez des exemples d'autres applications de cette théorie ?

Pour compenser ces sources d'erreurs, une nouvelle approche a été mise en place : les prévisions d'ensemble. Il est ici question de faire varier les conditions initiales dans le domaine d'incertitude des mesures et des modèles disponibles afin de pouvoir, comme expliqué sur Wikipédia : “permettre de quantifier la prédictibilité de l'atmosphère et d'offrir une marge d'erreur statistique sur la prévision.”

Pour infos, trois chercheurs, dont un très connus : Charney, Fjortoft et von Neumann réussirent la première prévision numérique du temps par ordinateur ¹⁴ . Les premiers programmes de prévisions numériques opérationnelles furent instaurés au début des années 1960.

La cosmologie

Prenons encore un autre domaine, et je pense que j'aurais bien illustré la situation : la cosmologie. Ici le but est de comprendre l'évolution des galaxies, la mécanique céleste, etc. Un bon exemple est de comprendre la forme des galaxies et notamment ce qu'il se passe quand deux galaxies se rencontrent, comme Andromède et la Voie Lactée vont le faire. A part la simulation, la seule solution que j'entrevois est d'attendre 4,5 milliards d'années et de prendre des photos pour voir ce que cela va donner ... Problème encore plus complexe en lien avec la cosmologie : Quelle est la répartition de masse dans l'univers et comment celle-ci a influencé son évolution jusqu'à nos jours ? La grande simulation Deus ¹⁵ qui a été mise en place par JM Alimi et son équipe au sein de l'Observatoire de Paris a, pour comprendre, tenté de simuler tout cela de manière massive sur le plus gros super ordinateur public de France nommé CURIE. La machine dispose de 92000 coeurs et sa performance crête est de 2 millions de milliards d'opérations à la seconde ¹⁶ . L'idée était ici de simuler tout l'univers depuis le Big Bang jusqu'à maintenant avec 3 scénarios différents de matière noire ^{17 18} :

¹⁴http://fr.wikipedia.org/wiki/Pr%C3%A9vision_num%C3%A9rique_du_temps

¹⁵<http://www.deus-consortium.org/>

¹⁶Pour comparer la machine la plus puissante qui est en Chine se nomme Tianhe-2. Elle possède un peu plus de 3 millions de coeurs et possède une performance de plus 33 millions de milliards d'opérations à la seconde.

¹⁷http://www.asa2012.phys.unsw.edu.au/HWWS/Harley%20Wood%20Winter%20School%202012/speaker%20notes/Davis_notes.pdf

¹⁸<http://www.deus-consortium.org/a-propos/dark-energy-universe-simulation-full-universe-run/cosmological-models-2/>

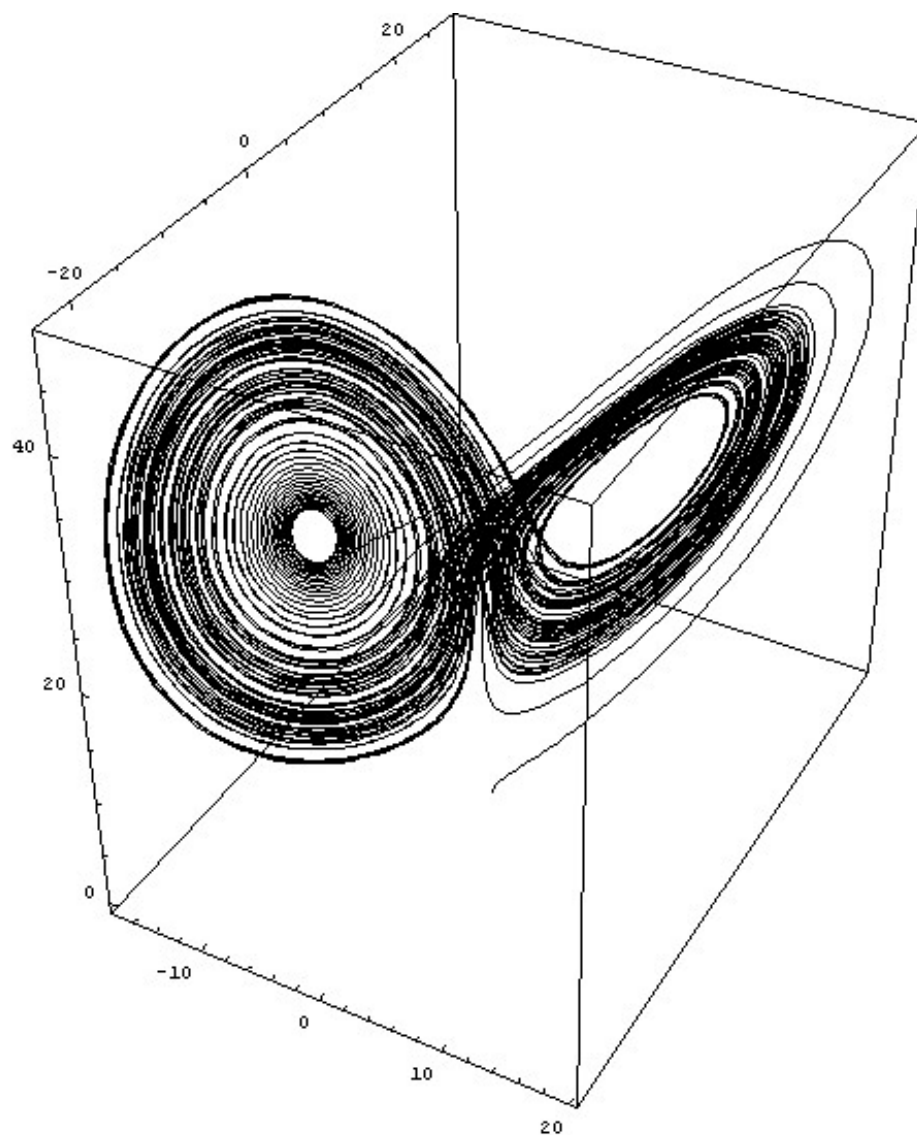


Figure 4: Attracteur étrange de Lorenz (1963) <http://commons.wikimedia.org/wiki/File:Lorenz.jpg>

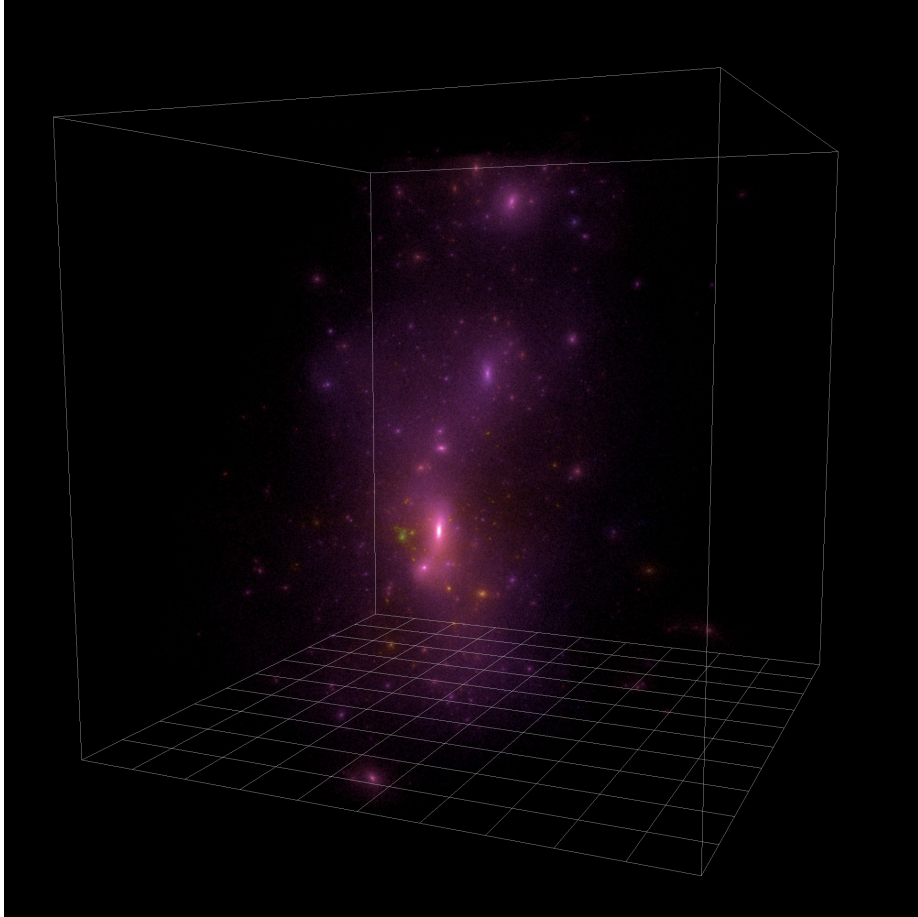


Figure 5: Halos pour une simulation 162 Mpc/h pour une cosmologie de type Λ CDM $w=5$ <http://www.deus-consortium.org/gallery/images/>

- Λ CDM : où l'on considère la relativité générale avec la constante cosmologique (équation d'état cosmologique ¹⁹ constante, pour simplifier le ratio nommé w entre la pression et la densité d'un fluide considéré parfait qui remplirait l'univers vaut -1) et de la matière noire froide (La matière noire froide ou **CDM**, de l'anglais *Cold dark matter*, est une théorie améliorée à partir de celle du Big bang. Elle repose sur le postulat supplémentaire que la majeure partie de la matière de l'Univers serait constituée d'éléments non observables par le rayonnement électromagnétique - noire - et dont les constituants se déplacent lentement - froide - ²⁰)
- modèle à base de quintessence ²¹: où l'on considère la relativité générale avec une énergie noire plus générique (qui évolue au cours du temps et qui correspond à un champ scalaire avec une équation d'état qui est dynamique, pour simplifier w varie au cours du temps) et de la matière noire froide
- modèle d'énergie fantôme ²²: qui est un dérivé du modèle à base de quintessence où l'équation d'état est satisfaite avec une valeur de w inférieure à -1. Ce modèle pousse notamment pour un univers aboutissant au "Big Rip" ²³.

Cela représentait la simulation de l'évolution de plus de 550 milliards de particules en utilisant 76000 coeurs de processeurs et plus de 300 Téra-octets (37500 DVDs) de mémoire vive pendant l'équivalent de 50 millions d'heures de calcul (presque 5700 ans de calcul en 28 jours)! Au final ils ont généré 1,5 Péta-octets de données utiles (187500 DVDs de données) qui serviront à ALMA, au VLT, au futur télescope spatial Euclid, mais aussi au Radio télescope Square Kilometer Array qui sera opérationnel en 2020. Pour ceux que cela intéresse, les données sont disponibles à l'adresse suivante : <http://www.deus-consortium.org/deuvo/>

Comme expliqué dans l'annonce de l'obtention du prix Joseph Fourier²⁴ 2012 pour le consortium DEUS ²⁵ :

Les détails du calcul, notamment le nombre et la taille des plus gros amas de matière existants, ne peuvent être vérifiés actuellement par

¹⁹[http://en.wikipedia.org/wiki/Equation_of_state_\(cosmology\)](http://en.wikipedia.org/wiki/Equation_of_state_(cosmology))

²⁰http://fr.wikipedia.org/wiki/Mati%C3%A8re_noire_froide

²¹[http://en.wikipedia.org/wiki/Quintessence_\(physics\)](http://en.wikipedia.org/wiki/Quintessence_(physics))

²²http://en.wikipedia.org/wiki/Phantom_energy

²³<http://www.fayard.fr/desir-dinfini-9782213635118>

²⁴Ce prix récompense une personne ou une équipe pour leurs travaux dans le domaine de la parallélisation des applications de simulation numérique sur des architectures traditionnelles ou hybrides, réalisés dans le cadre d'un laboratoire français, public ou privé. Bull et GENCI (Grand Équipement National de Calcul Intensif, structure privée ayant pour rôle de fédérer les actions de développement de la simulation et du calcul intensif en France) souhaitent ainsi favoriser le développement et l'amélioration de la simulation numérique en France, tant le domaine de la recherche scientifique que dans toutes ses applications industrielles. Ce prix rend hommage à Joseph Fourier dont les travaux ont largement contribué à la modélisation mathématique des phénomènes physiques.

²⁵<http://www.obspm.fr/prix-joseph-fourier-2012-a-l-equipe-dark-energy.html>

les observations au télescope. Pour cela, il faudra attendre la mission de cosmologie spatiale Euclid de l'ESA, au lancement en 2019.

D'autres domaines où les limitations ne sont pas forcément physiques

On pourrait illustrer cela avec encore d'autres domaines comme la génétique où l'on ne peut pas tester toutes les maladies avec tous les remèdes possibles pour d'évidentes raisons éthiques ... Ou bien la physique nucléaire où soit il est nécessaire d'avoir des instruments très complexes à disposition pour faire des expériences (LHC, Iter, etc), soit il faut pouvoir faire des tirs de bombes nucléaires pour observer ce qu'il se passe (mais ce n'est plus autorisé dans nos démocraties, et puis de manière générale ce n'est pas top).

Un autre domaine où il est compliqué aussi de faire des expériences en grandeur réelle est celle de la sélection naturelle et/ou de l'évolution biologique, le podcast de Marco sur l'expérience de Lenski montre bien que c'est complexe et que la simulation peut avoir un vrai apport. David pourra sûrement nous en parler plus en détail!

David, est ce que tu as des précisions sur les aspects de simulation de l'évolution biologique et de la sélection naturelle ?

On pourrait aussi parlé du Humain Brain Project, un des grands projets stratégique de l'Europe où il est question de simuler le cerveau humain ²⁶.

Attention

Alors attention, je ne prétends pas que la simulation est la réponse ultime et qu'elle permet de tout prendre en compte et tout comprendre. Tous les phénomènes physiques ne pourront pas forcément être pris en compte (comme expliqué pour la simulation de la météorologie), et les modèles utilisés ... sont des modèles. Ils intègrent donc des hypothèses qui ne seront pas forcément vraie IRL ou des simplifications qui permettent à la simulation d'être pertinente en temps de calcul ou d'espace mémoire à utiliser.

Et comme je l'ai dit, la simulation doit être validée pour certains cas, afin de s'assurer que le modèle utilisé pour représenter le phénomène étudié est valable et que la simulation numérique basée sur ce modèle est pertinente (les techniques de discrétisation et de calcul doivent aussi être validée). Et pour cela l'expérimentation ou tout du moins l'observation est indispensable.

Problèmes des modèles

Il est important de noter que l'on ne pourra rien apprendre sur ce qui n'est pas inclus dans notre modèle. On ne peut pas faire de découverte. Il ne faut pas

²⁶<https://www.humanbrainproject.eu/fr>

faire croire que l'on peut se servir des modèles pour découvrir des phénomènes, au contraire, pour des modèles complexes, on peut en arriver à des résultats qui peuvent être totalement inattendus : non pas que ce soient des découvertes, mais des conséquences d'interactions des modèles multi-échelle (comme pour la simulation un système biologique au niveau macro-cellulaire, micro-cellulaire, biochimique, génétique, etc) et multi-physique (comme pour la simulation de mécanique des fluides en conditions extrêmes : air/eau/glace résistance des matériaux, thermodynamique, etc) entre leurs composantes.

Encore un point sur les modèles : les modélisation choisies pour représenter des phénomènes sont choisies la plupart de temps pour étudier certains phénomènes. Quand on veut étudier le comportement d'une aile d'avion au sein d'un écoulement d'air, on peut proposer des modèles différents si l'on veut étudier :

- son comportement avec un écoulement sub-sonique
- son comportement avec un écoulement super-sonique
- son comportement dans des conditions extrêmes (air + basse température + eau)

Pour chacune de ces simulations, on va avoir des équations qui seront modélisées de manière différente car les composantes qui seront dimensionnantes seront différentes, voire il sera nécessaire d'ajouter des éléments à la simulation comme la prise en compte de la température, où la prise en compte des interactions air/eau, etc.

Première simulation : Méthode de monte carlo et physique nucléaire

Pour finir ce podcast, je voulais parler de la première simulation numérique qui fut réalisée. Cette simulation numérique a été effectuée en 1953 sous l'impulsion de John Van Neumann dans le domaine de la physique nucléaire ^{27 28 29}. Il souhaitait pouvoir, dans le cadre du projet Manhattan, simuler le comportement des neutrons lors des réactions en chaîne dans des armes explosives basées sur la fission. Il est apparu qu'une bonne méthode pour arriver à faire cela est celle de Monté Carlo. Comme le précise Wikipédia, "Le nom de ces méthodes, qui fait allusion aux jeux de hasard pratiqués à Monte-Carlo, a été inventé en 1947 par Nicholas Metropolis, et publié pour la première fois en 1949 dans un article coécrit avec Stanislaw Ulam". S. Ulam que l'on retrouve notamment dans l'expérience de John Van Neumann qui s'appelle d'ailleurs l'expérience de Fermi-Pasta-Ulam (Fermi était le chercheur, Pasta l'informaticien et Ulam le mathématicien).

²⁷http://fr.wikipedia.org/wiki/Exp%C3%A9rience_de_Fermi-Pasta-Ulam

²⁸<http://132.187.98.10:8080/encyclopedia/en/monteCarlo.pdf>

²⁹http://fr.wikipedia.org/wiki/M%C3%A9thode_de_Monte-Carlo

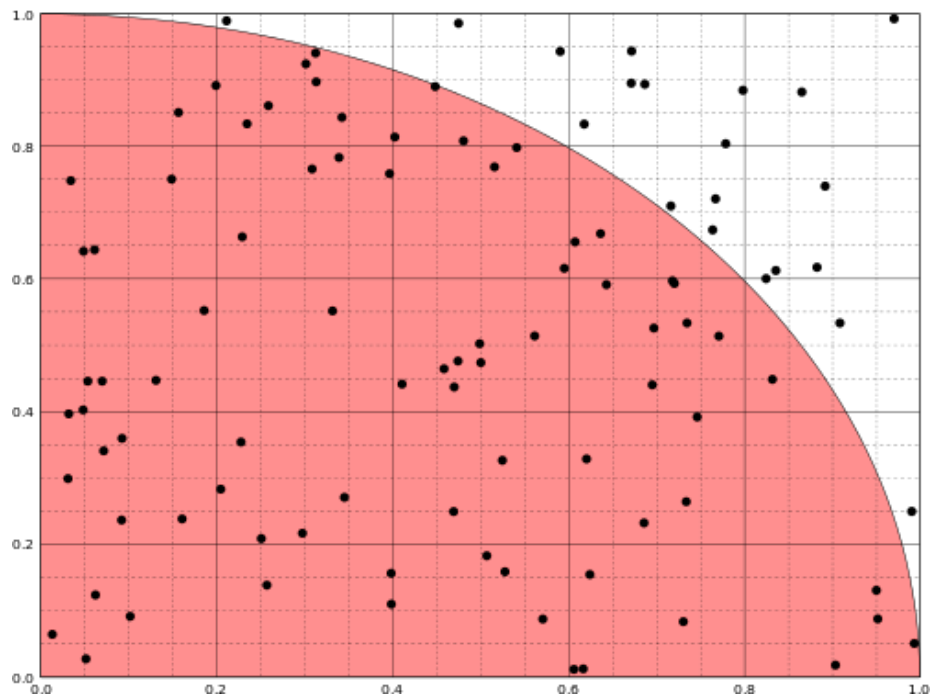


Figure 6: Représentation du calcul de la valeur de pi par rapport au nombre de points aléatoires étant contenus dans un quart de cercle, l'ensemble des possible étant un carré de côté R. L'exemple est pris avec 100 points <http://commons.wikimedia.org/wiki/File:Montecarlo-valeur-pi.svg>

La méthode proposée (qui est la base de toute bonne méthode de Monte Carlo) consiste dans le découpage d'un processus complexe en un grand nombre de petites étapes, où, à chaque pas plusieurs choix sont possibles avec pour chacun d'entre eux une probabilité définie de se réaliser. Chaque run complet va fournir un résultat et en effectuant un grand nombre on sera en capacité de connaître la fonction de probabilité du processus complexe global en incluant l'incertitude inhérente due au hasard.

John Von Neumann a immédiatement reconnu les avantages de cette méthode et notamment dans son application à l'exploration du comportement des chaînes de réaction à base de neutrons. Et ceci notamment dans les armes basées sur la fission, pour la prédiction de leurs explosions, mais aussi pour les problèmes de diffusion et de multiplication des neutrons.

Dans la simulation que faisait John Von Neumann il y avait plusieurs étapes qui peuvent être représentées de manière grossière sous forme de diagramme :

- La première était la génération d'un neutron dont la position et la vitesse était fournies de manière aléatoire. Les lois de probabilité correspondantes sont déterminées afin de refléter les conditions initiales réputées connues.
- La seconde étape consistait en un déplacement libre du neutron sur une certaine distance définie par une distribution de probabilité spécifique qui va dépendre par les propriétés du matériaux et les sorties de la première étape.
- Trois cas peuvent ensuite apparaître pour l'étape suivante :
 - Soit le neutron entre en collision au sein du même matériaux
 - Soit le neutron traverse la frontière afin de réaliser un nouveau déplacement libre dans ce nouveau matériau.
 - Soit la branche se termine car le neutron a été absorbé par le matériau sans générer de nouveau neutron ou est sorti complètement de l'objet.

On réalisait ensuite un grand nombre de ces enchaînements et on observait les résultats pour arriver à comprendre le fonctionnement des réactions en chaîne que ces opérations visaient à simuler.

Ainsi pour chacune de ces étapes, des distributions de probabilités sont utilisées pour chacun des cas identifiés et il est donc nécessaire de disposer d'une bonne source de nombres aléatoires.

Importance de la qualité du hasard

Ce que l'on remarque et qui a déjà été abordé dans de précédents podcasts (notamment ceux de Robin sur le hasard), c'est qu'il est important d'avoir une bonne source de nombres aléatoires.

Il est tout à fait possible d'utiliser des générateurs de nombres aléatoires qui sont basés sur des phénomènes assez imprévisibles pour être considérés aléatoires (dés, roulette, etc). Le problème est que le temps d'obtention des nombres peut-être lent, voire trop lent pour les expériences à réaliser pour lesquels il faut parfois un grand nombre ... de nombres aléatoires ³⁰ .

Il existe aussi des générateurs de nombres aléatoires reposant sur des phénomènes physiques (radioactivité, bruits thermiques, mécanique quantique, etc).

Il existe enfin des algorithmes qui peuvent être utilisés pour générer des nombres aléatoires. Nous sommes tous d'accord que des algorithmes déterministes, exécutés sur des machines déterministes (les ordinateurs), ne risquent pas de donner des nombres vraiment aléatoires. C'est d'ailleurs pour cela que l'on parle plutôt de nombre pseudo-aléatoires ³¹ . L'intérêt de ces générateurs de nombres pseudo-aléatoires (ou Pseudo-Random Number Generator, PRNG), est qu'ils permettent de fournir des nombres qui présentent certaines propriétés du hasard, comme par exemple l'indépendance des nombres entre eux ou le fait qu'il est difficile de repérer des groupes de nombres suivant certaines règles. L'autre intérêt est tout simple : ils peuvent être implémentés et peuvent générer des nombres très rapidement sans avoir de matériel spécifique (comme pour les générateurs physiques). Il existe d'ailleurs plusieurs ensembles de suite de tests qui ont été mises en place pour valider le hasard produit par ces PRNG :

- La DieHard test suite de George Marsaglia (<http://stat.fsu.edu/pub/diehard/>) comme le film avec notre Bruce Willis préféré.
- La TestU01 de Pierre L'Ecuyer et Richard Simard (<http://www.iro.umontreal.ca/~simardr/testu01/tu01.html>)

Le but de ces tests est de confronter les PRNG à des situations extrêmes dans lesquels l'aspect "non-aléatoire" peut apparaître, comme des corrélations entre des suites de nombres fournis par ces générateurs. Ces tests vont par exemple tester la répartition dans un cube des nombres générés quand on en prend trois consécutifs pour faire une coordonnée ou encore le test dit de "parking lot" en plaçant des voitures sur une parking de carré 1 (chaque voiture est une sphère). Si il y a des crashes (une voiture veut se garer sur une place déjà prise), on recommence. Au bout d'un certain nombre de tentatives, on regarde le nombre de voitures garées, et on doit s'attendre à obtenir une probabilité qui suit une loi normale. On fait ce que l'on appelle un test dit d'adéquation à la loi normale : on vérifie la compatibilité des données avec une distribution de type loi normale. Wikipédia en anglais possède une page dédiée aux tests de DieHard : http://en.wikipedia.org/wiki/Diehard_tests

³⁰http://fr.wikipedia.org/wiki/G%C3%A9n%C3%A9rateur_de_nombres_al%C3%A9atoires

³¹http://fr.wikipedia.org/wiki/G%C3%A9n%C3%A9rateur_de_nombres_pseudo-al%C3%A9atoires

Parmi les failles qu'il faut éviter chez ces PRNG, on peut notamment citer :

- La variation de la qualité du hasard des nombres générés en fonction de la graine fournie
- Valeurs successives qui ne sont pas indépendantes
- Une période trop petite
- Certains bits qui ne sont pas assez aléatoire dans les sorties (en informatique toutes les données sont représentées sous forme de 0 et de 1 et assemblés en un certain nombre d'octets (ensemble de 8 bits). Les entiers sont par exemple codés sur 32 bits, ou 4 octets)

Parmi les PRNG les plus connus, Robin avait déjà parlé de celui de John Van Neumann (Robin avait notamment abordé ses failles avec l'histoire des zéros).

Robin tu pourrais juste nous redonner l'explication du fonctionnement de ce générateur s'il te plaît ?

Avec Nico ils avaient abordé le cas des générateurs congruentiels linéaires ³² (à base de module) qui sont des suites numériques dites arithmético-géométriques et dont on prend le modulo d'une certaine valeur (la valeur à l'étape suivante est la somme du produit de la valeur à l'étape courante par un coefficient (a) plus un nombre (b) et l'on prend le module pour une certaine valeur (m)).

Ces générateurs ont été introduits par Lehmer en 1948 sous une forme réduite (b était nul). Ici la période est au maximum m (dans les implémentations 32 bits et bien c'est 2^{32}). Un des générateurs de ce type qui a été le plus utilisé se nomme RANDU ³³. Il fut introduit par IBM dans les années 60. Pour lui, les paramètres a et m valent respectivement 65539 et 2^{31} . Ce générateur a beaucoup été utilisé, jusqu'à ce que l'on se rende compte qu'il générât des nombres, qui utilisés pour représenter des points en dimension supérieur à 2, n'étaient pas placés au hasard dans les cubes (ou hypercubes en dimensions supérieures) considérés. Pour les plus ferrus de mathématiques, RANDU échoue violemment sur le test spectral, notamment à partir de la dimension 3.

Pour finir sur les PRNG, l'un des plus connus pour être de bonne qualité (sauf pour la crypto) est celui inventé par Makoto Matsumoto et Takuji Nishimura en 1997 nommé Mersenne Twister ³⁴. Il possède déjà une très grande période : $2^{19937} - 1$, il est uniformément distribué sur 623 dimensions (pour des nombres sur 32 bits), il est très rapide, aléatoire quelque soit le poids du bit considéré et pas les tests de Die Hard (dont je parle plus tôt). Je ne vais pas forcément rentrer dans les détails de son implémentation, bien plus complexe que les LCG, mais l'idée est de faire des déplacements de bits, et des opérations logiques de type OU exclusif sur les éléments d'une matrice qui contient des bits de nombres

³²http://fr.wikipedia.org/wiki/G%C3%A9n%C3%A9rateur_congruentiel_lin%C3%A9aire

³³<http://fr.wikipedia.org/wiki/RANDU>

³⁴http://fr.wikipedia.org/wiki/Mersenne_Twister

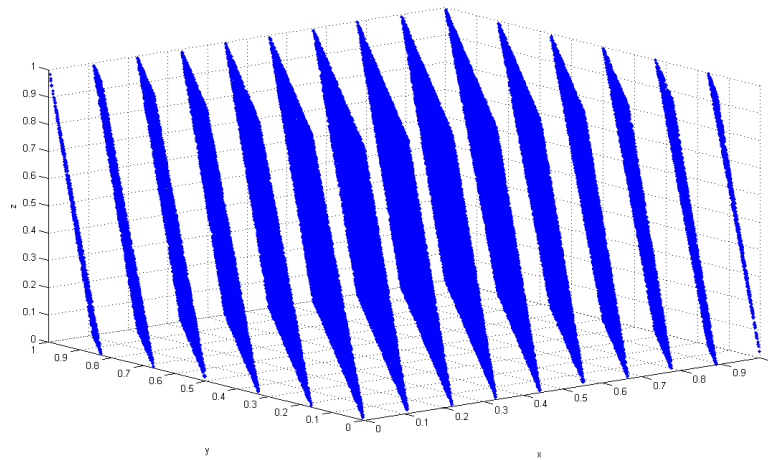


Figure 7: Représentation créée avec MATLAB en générant 100002 valeurs utilisant RANDU et en positionnant dans un graphe 100000 triplets (x,y,z) consécutifs <http://en.wikipedia.org/wiki/File:Randu.png>

aléatoires et qui est réutilisée, un peu comme avec les LCG pour l'itération suivante.

Un des derniers problèmes avec l'usage de PRNG est dans le cas de parallélisation de code. Quand un code est déployé sur des serveurs de calcul, une des difficultés qui apparaît est de pouvoir générer pour chaque bout de code une série de nombres aléatoires qui sont indépendants et identiquement distribués et que notamment, on ne se retrouve pas à utiliser les mêmes nombres pseudo-aléatoires sur chaque noeud de calcul.

Je pense que je vais m'arrêter là sur la simulation pour ce podcast. On a pu voir les raisons du besoin de simulation, des difficultés rencontrées avec différents champs de recherche et la première simulation numérique qui était probabiliste.

Dans la suite de ce podcast, on ira plus sur les simulations pour des équations déterministes. On verra trois grandes méthodes utilisées pour passer des modèles à l'ordinateur, et ensuite les ramifications en terme de mathématiques avec les méthodes utilisées pour résoudre les problèmes posés et ce que l'informatique a fait pour les résoudre dans des temps corrects.

Notes

Paramétrisation de maille en prévision numérique du temps

La représentation de l'influence moyenne à grande échelle des phénomènes de la petite échelle est appelée *paramétrisation*. Les phénomènes sous-maille les plus communément paramétrisés par les concepteurs des modèles sont :

- La convection verticale (dont font partie les orages)
- La physique des nuages (condensation, collection et collision des gouttes, effet Bergeron, changements de phase, etc.)
- Les effets radiatifs atmosphériques (rayonnement de la chaleur)
- L'interface surface-air :
- Échanges de chaleur et d'humidité entre la surface et l'atmosphère
- Frottement et turbulence près du sol
- L'effet des montagnes et des irrégularités du terrain :
- Effet de blocage du vent
- Ondes atmosphériques en aval des montagnes

Le paramétrage des phénomènes physiques ne compense pas complètement les limitations imposées par un espacement trop grand de la maille des modèles. Le choix et l'ajustement des schémas de paramétrisation a un impact important sur la qualité des prévisions.

Matière noire, énergie noire et la composition de l'univers

Si l'on souhaite s'intéresser à ces notions, il est possible de se référer aux sources suivantes :

- Wikipédia ³⁵ et ³⁶
- Le livre de Trinh Xuan Thuan : Désir d'infini ³⁷
- Ou encore le très bon dossier de Mathieu, l'épisode 31 qui traitait de la question ³⁸

En ce qui concerne la matière noire il y a grosso modo deux grands types de matière noire :

- La matière noire chaude : composée de particules très rapides dont la vitesse est proche de celle de la lumière

³⁵http://fr.wikipedia.org/wiki/Mati%C3%A8re_noire

³⁶http://fr.wikipedia.org/wiki/%C3%89nergie_sombre

³⁷<http://www.fayard.fr/desir-dinfini-9782213635118>

³⁸<http://www.podcastscience.fm/dossiers/2011/04/07/dossier-la-matiere-noire/>

- La matière noire froide : composée de particules plus massives et plus lentes

Comme expliqué sur Wikipédia et très bien aussi dans le dossier de Mathieu :

- Si l’Univers était dominé par de la matière noire chaude, la très grande vitesse des particules la constituant empêcherait dans un premier temps la formation d’une structure plus petite que le superamas de galaxies qui ensuite se fragmente en amas de galaxies, puis en galaxies, etc. C’est le scénario dit “du haut vers le bas”, puisque les plus grosses structures se forment d’abord, pour ensuite se diviser. Le meilleur candidat pour constituer la matière noire chaude est le neutrino.
- En revanche, si la matière noire froide dominait l’Univers, les particules vont parcourir une distance plus petite et donc effacer les fluctuations de densité sur des étendues plus petites que dans le cas de matière noire chaude. La matière ordinaire va alors se regrouper pour former d’abord des galaxies (à partir de nuages de gaz), qui elles-mêmes se regrouperont en amas, puis superamas. C’est le scénario dit “du bas vers le haut”. Les candidats à la matière noire froide sont les WIMP et les MACHO.”

Actuellement, c’est le modèle de matière noire froide qui semble l’emporter. En effet, les galaxies sont en équilibre dynamique, ce qui laisse penser qu’elles se sont créées avant les amas — dont tous ne semblent pas encore stables — à qui il faut plus de temps pour atteindre cet équilibre. Cependant, les théories introduisent aujourd’hui un peu de matière noire chaude. Celle-ci est nécessaire pour expliquer la formation des amas ; la matière froide seule ne pouvant la permettre en si peu de temps.

Énergie fantôme

Comme expliqué sur Wikipédia ³⁹ :

En cosmologie, l’énergie fantôme désigne une forme hypothétique d’énergie dont la densité aurait la particularité surprenante d’augmenter lors de l’expansion de l’Univers. Cette énergie est un candidat potentiel à l’énergie noire. Cette forme d’énergie noire implique la violation d’un principe physique connu sous le nom de condition faible sur l’énergie ⁴⁰ (valable pour les fluides parfaits). Cet argument est l’un des principaux écueils théoriques à l’existence de l’énergie fantôme. Cependant, la condition faible sur l’énergie n’est valable que pour les fluides parfaits. L’énergie sombre pourrait donc ne pas enfreindre ce principe, à condition de postuler une interaction avec d’autres composants cosmologiques, comme, par exemple, la matière sombre. L’énergie fantôme, si elle existe, serait responsable d’un emballement de l’expansion de l’Univers, qui causerait un éloignement

³⁹http://fr.wikipedia.org/wiki/%C3%89nergie_fant%C3%B4me

⁴⁰http://fr.wikipedia.org/wiki/Condition_sur_l%27%C3%A9nergie

arbitrairement grand des différents objets célestes les uns des autres en un temps fini, puis une dislocation de ceux-ci : c'est le modèle cosmologique du Big Rip⁴¹. Dans ce scénario, l'énergie noire prend le pas sur la gravité de manière exponentielle et la densité de l'univers atteint une valeur infinie en un temps fini. La force devient tellement grande qu'elle surpasse les forces électromagnétiques, faible et forte et toute la matière se décompose et se détruit jusqu'à ses composants les plus fondamentaux qui ne peuvent plus rester unis.

Ce modèle diffère du Big Crunch, dans lequel l'expansion s'arrête pour laisser place à une phase de contraction. Dans ce dernier cas, les objets astrophysiques sont détruits car ils entrent en collision les uns avec les autres (ils sont "écrasés", d'où Big "Crunch"), alors que dans le cas du Big Rip, ils sont étirés par une expansion de plus en plus violente, jusqu'à être disloqués (ils sont « déchirés », d'où Big "Rip"). L'hypothèse du scénario du Big Rip est pour l'heure marginalement compatible avec les données observationnelles et reste donc relativement spéculative.

⁴¹http://fr.wikipedia.org/wiki/Big_Rip