

PROSPECTIVE GRILLE DE PRODUCTION

Sciences de la Planète, de l'Univers et de l'Environnement

M. Petitdidier (IPSL), F. Le Petit (Obs. Paris),
G. Moguilny (IPGP), P. Le Sidaner (Obs. Paris),
S. Godin-Beekmann (IPSL), K. Ramage (IPSL),
S. Manoussis (CNRS/INSU)

Sciences de la Planète, de l'Univers et de l'Environnement

- **Mosaïque de disciplines et sous-disciplines**

- Participation à des Programmes de Recherche régionaux, nationaux, Européens, internationaux
- Expérimentation
 - Grande variété d'instrumentations: Thermomètre,.... Radar
 - Données : différentes résolutions (1ms à 1 point/par semaine), volumes, formats standards
- Modélisation
 - Grande variété : 1 CPU jusqu'à 2000 CPUs (Earth Simulator)
 - Temps de calcul : quelques minutes à plusieurs mois

- **Besoins en moyens de calcul**

- Super Calculateurs
- Stations de travail
- Clusters

- **Organisation du stockage**

- Mesocentres existants ou en projet
- Stockage local ou dans les Centres de données

Prospectives

- 3 prospectives menées en parallèle
 - Super calculateurs
 - Mesocentres
 - Grille de production/domaine
- Outils pour la prospective “Grille” Planète, Univers, Environnement
 - **Réunion** d’information 27 mai à l’observatoire de Paris :
 - Présentation de la grille EGEE, des applications en Sciences de la Planète et Univers, utilisation et accès à EGEE, EGEEIII, Interfaces (Geocluster, Ganga, gEclipse) et prospectives du ministère et des universités
 - ~30 représentants de différentes disciplines en Ile de France
 - **Questionnaire** web sur limesurvey avec des liens pour les présentations sur EGEE et les applications en Sciences de la planète et de l’Univers, et le document de prospective de l’IdG

Diffusion du questionnaire

- Tous les laboratoires et équipes relevant de Planète Univers avec un mot de S. Manoussis
- Listes de diffusion
 - Thématiques (Radar Strato-Troposphérique, Geocluster, Société Française d'A&A, Observatoire Virtuel...)
 - Instituts (IPSL, IPGP, Observatoires.....)
- Organismes contactés
 - BRGM : équipe déjà travaillant sur EGEE
 - CEMAGREF : Directeur d'Antony (IUGG), responsable
 - CNES : Groupe "Grille"; CNES/INSU (Postel, ICARE et Ether)
 - IFREMER : Individus
 - Météo France : Directeur CNRM, équipe "Grille"
- Contacts individuels
 - Visiteurs stand EGEE aux Conférences IUGG, EGU et ENVISAT
 - Personnes d'autres laboratoires avec qui on est en contact

Résultats

- 162 réponses dont 85 exploitables
 - Certaines sont trop incomplètes pour être exploitées
 - Un certain nombre d'individus ont répondu au nom d'une équipe ou d'un laboratoire.
L'équipe pouvant aller de 3 personnes à 50 personnes
 - Des utilisateurs d'EGEE n'ont pas répondu
 - Peu de participants à la réunion d'information du 27 mai ont répondu
 - Des équipes et laboratoires ne se sont pas sentis concernés ou n'ont pas été contactés via les différentes listes
 - Beaucoup ne connaissait pas la grille
- Intéressant car des équipes ont répondu être intéressées sans avoir été en contact précédemment avec des personnes travaillant sur la grille

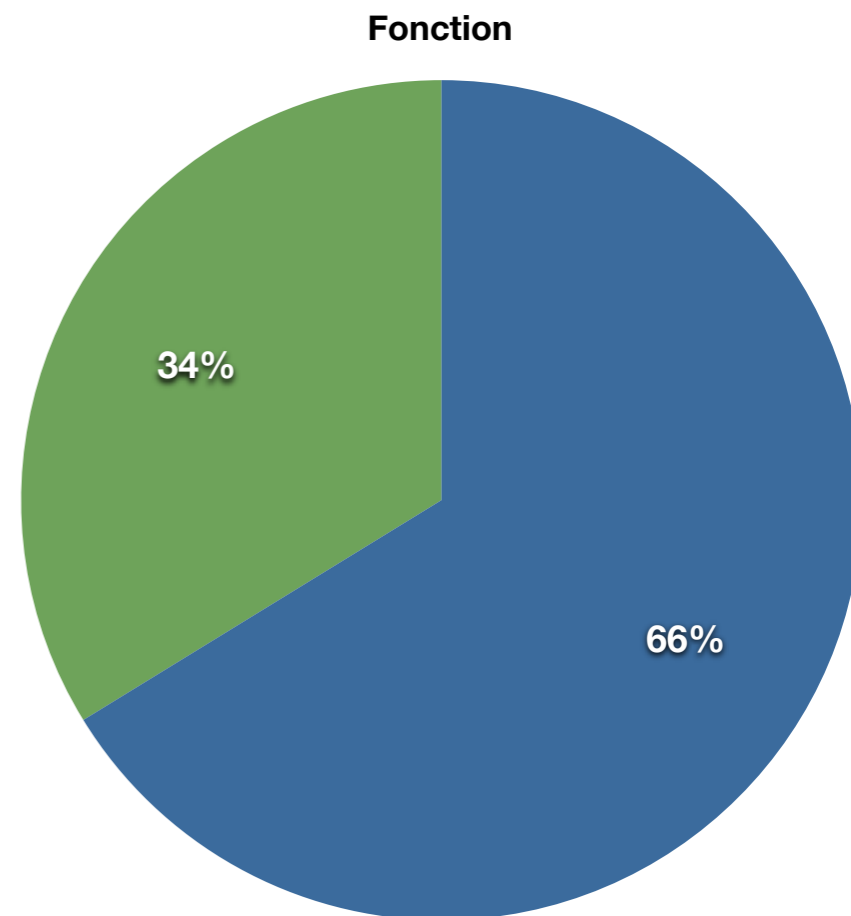
Répartition régionale

Ile de France	IPSL, IPGP, Obs. de Paris, CEA, Univ. Paris Sud, LISA, BRGM, IGN, UPMC/ tectonique	39
PACA	Geosciences-azur, Obs. Côte d'Azur, Montpellier (Geosciences), GRAAL, LAM, CPPM, CEREGE	13
Rhône-Alpes	Univ. Savoie (Astro. et ES), UJF(LGGE, LGIT, Planétologie, Astrophysique), Obs. Besançon	13
Midi-Pyrénées	Obs. Midi-Pyrénées, Meteo France, LATT, Praksys, LGEI/Ecole des mines d'Alès	8
Bretagne	Univ. de Nantes, Univ. de Rennes, Univ. de Brest	4
Auvergne	OPGC	2
Est	CDS, EOST/Strasbourg	2
Nord	LOA	2

Statistiques sur les personnes ayant répondues

Fonction

	Nombre	%
Individu	51	66,2
Responsable d'équipes et directeurs	26	33,8
Total	77	100,0

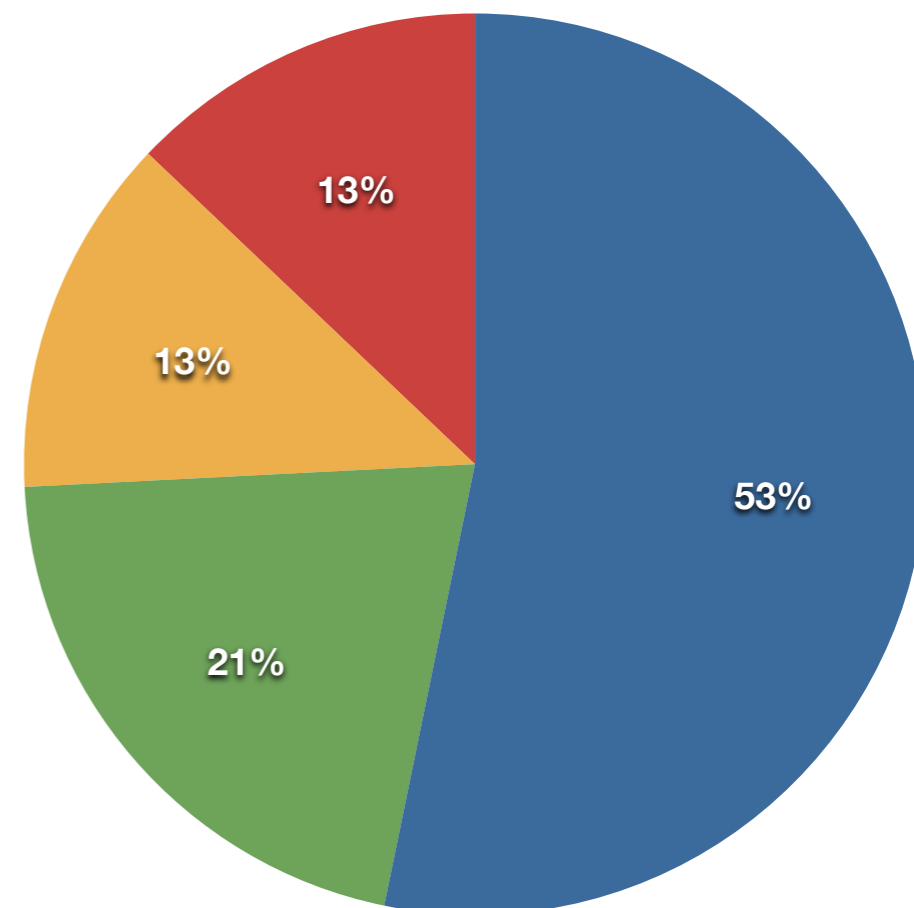


- Individu
- Responsable d'équipes et directeurs

Statuts

	Nombre	%
Chercheur permanent	33	53,2
Ingénieur	13	21,0
Doctorant	8	12,9
Autres (post-docs, CDD, ...)	8	12,9
Total	62	100,0

Statuts



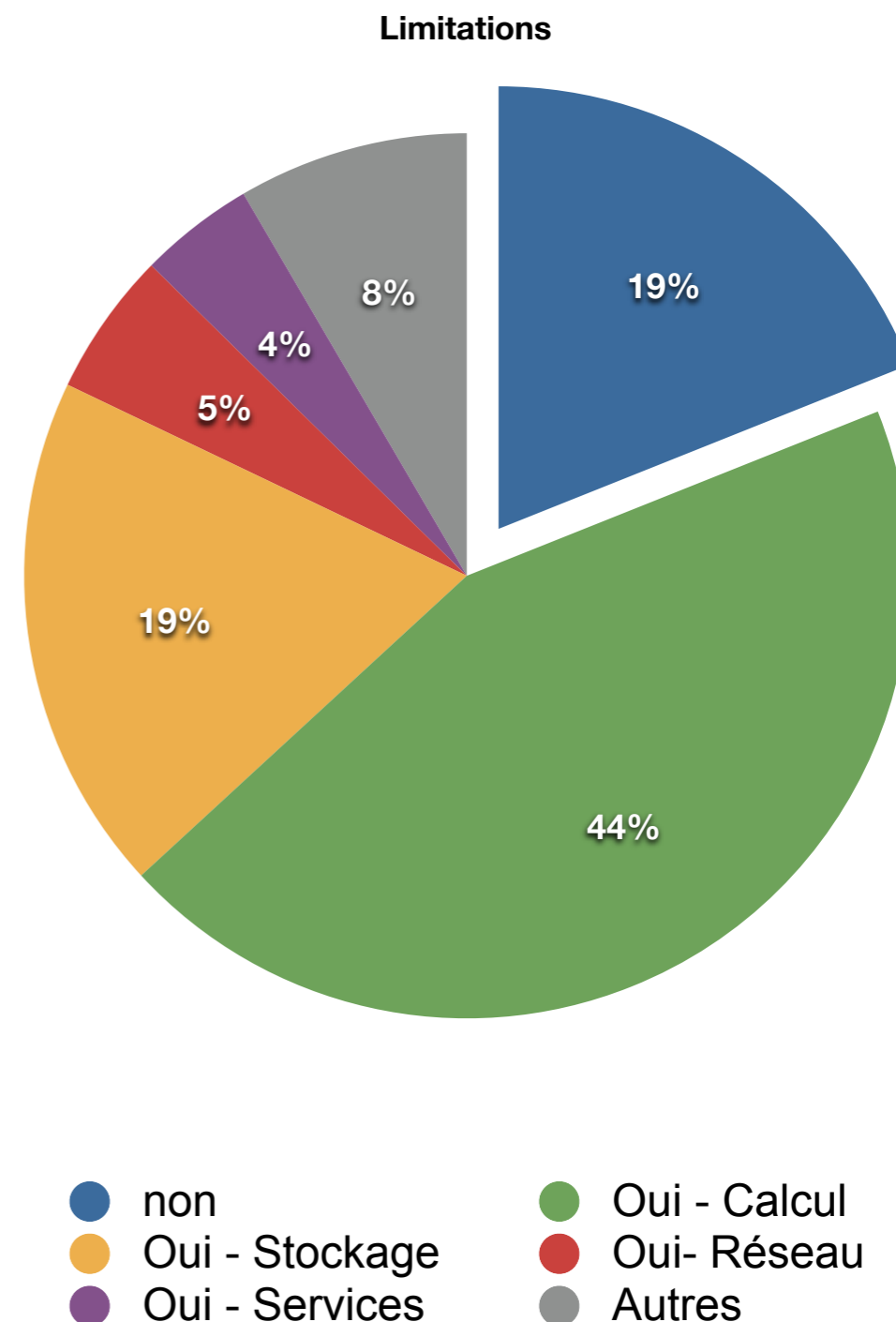
- Chercheur permanent
- Ingénieur
- Doctorant
- Autres

Limitations à cause des ressources informatiques

Etes-vous **limité par les ressources informatiques** actuellement à votre disposition pour mener à bien vos recherches ?

	Nombre	%
non	18	18,9
Oui - Calcul	42	44,2
Oui - Stockage	18	18,9
Oui - Réseau	5	5,3
Oui - Services	4	4,2
Autres	8	8,4
Total	95	100,0

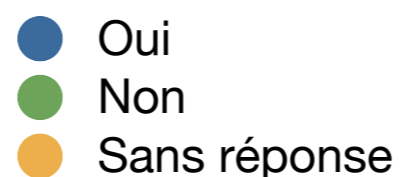
Limitations mentionnées : **moyens humains**



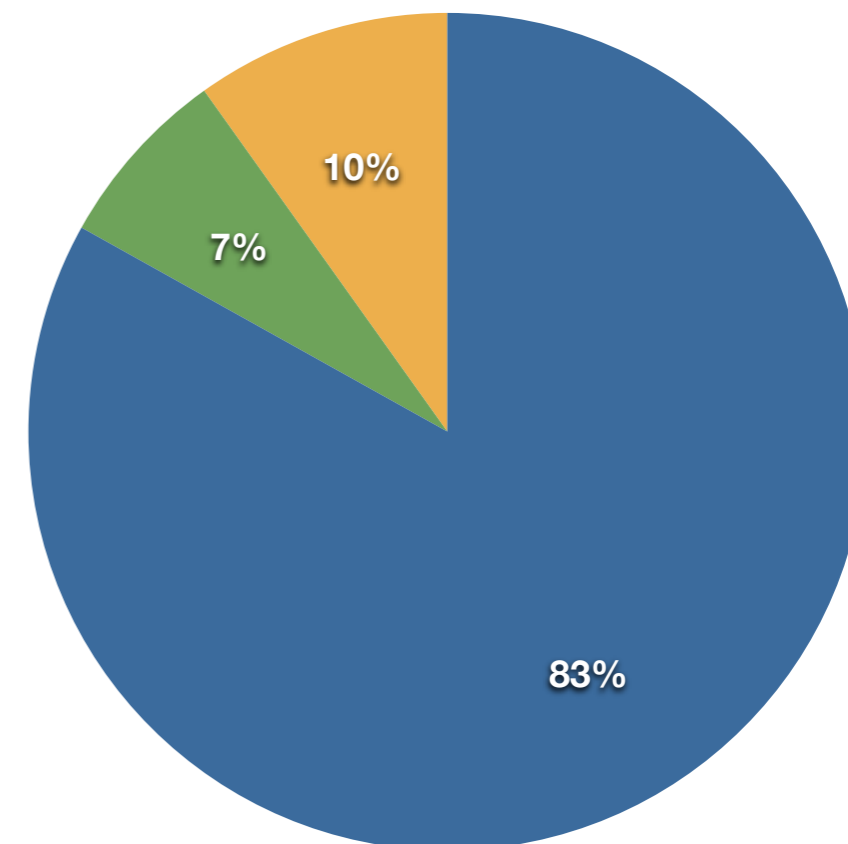
Applications répétitives

Travaillez-vous avec des applications dans lesquelles vous devez **exécuter un grand nombre de fois un programme** en faisant varier ses paramètres ?

	Nombre	%
Oui	59	83,1
Non	5	7,0
Sans réponse	7	9,9
Total	71	100,0



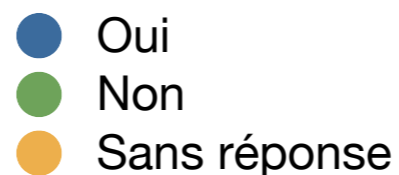
Applications répétitives



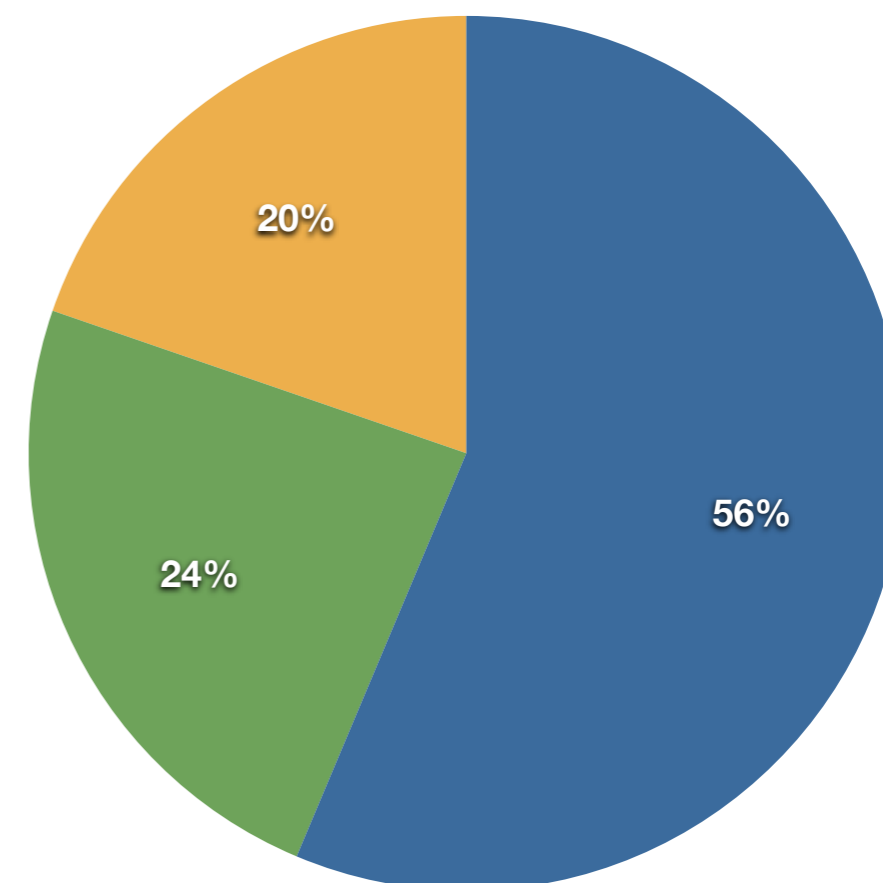
Possibilité de répartir les applications

Dans le cas d'un besoin en temps CPU très important, est-il possible de **répartir sur plusieurs machines indépendantes l'application** que vous traitez ou voudriez traiter ?

	Nombre	%
Oui	40	56,3
Non	17	23,9
Sans réponse	14	19,7
Total	71	100,0



Répartition des applications



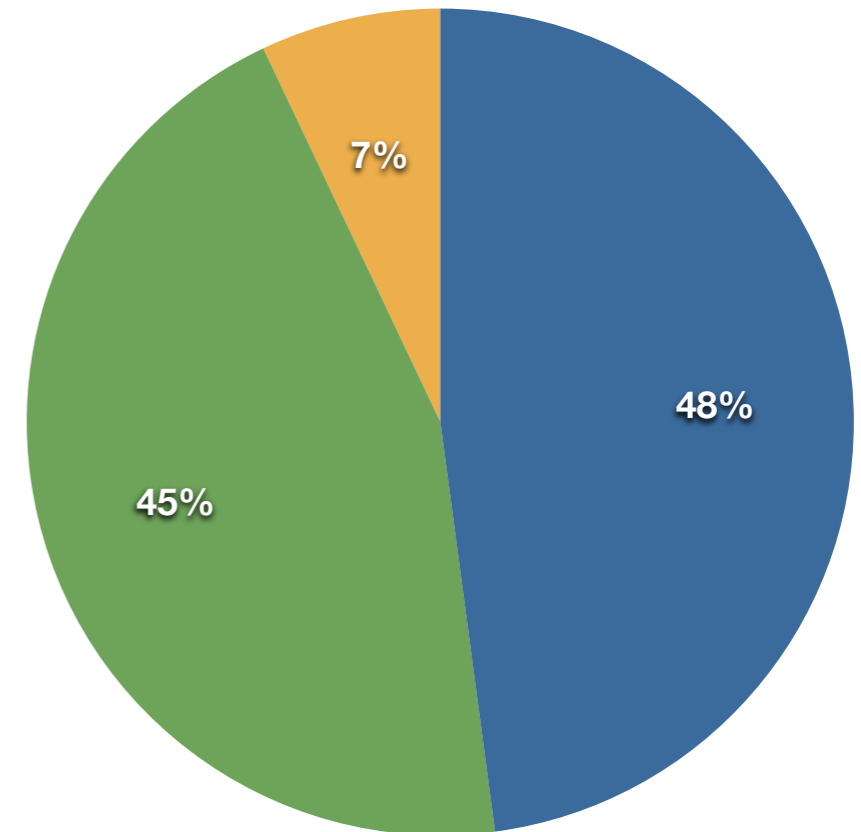
Parallélisation des applications

Vos applications sont-elles **parallélisées** ?

	Nombre	%
Oui	34	47,9
Non	32	45,1
Sans réponse	5	7,0
Total	71	100,0

- Oui
- Non
- Sans réponse

Applications parallélisées

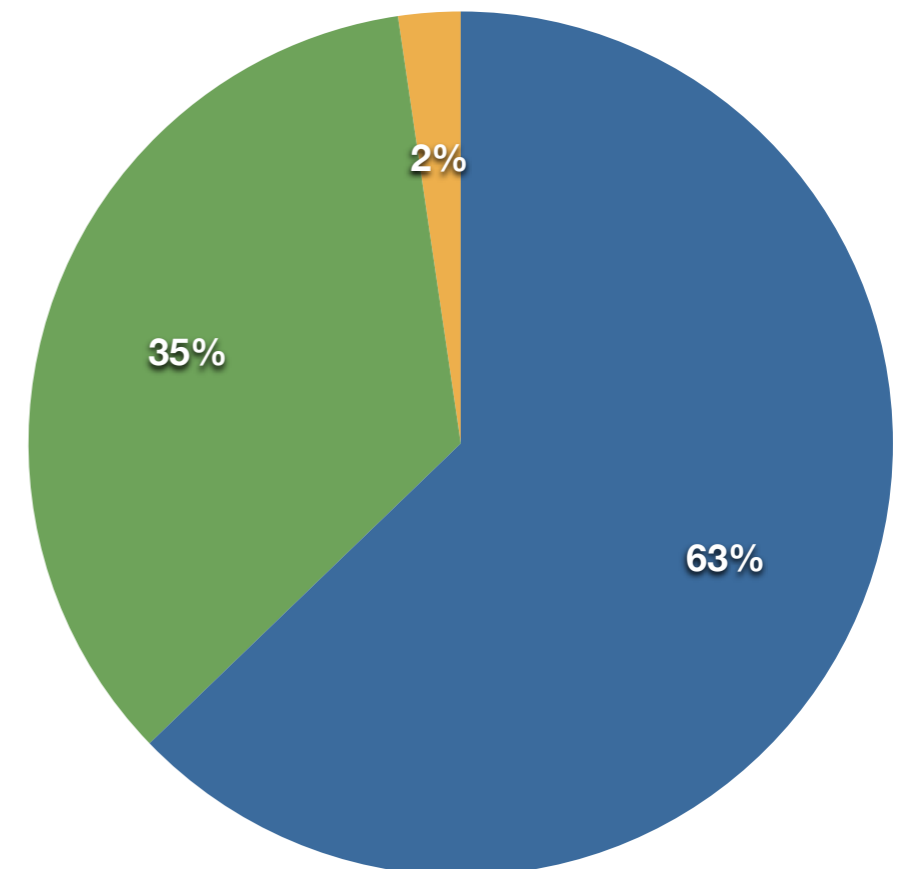


Quel **type** de parallélisme ?

	Nombre	%
MPI	27	62,8
OpenMP	15	34,9
Autre	1	2,3
Total	43	100,0

- MPI
- OpenMP
- Autre

Type de parallélisme



Autre: Embarassing parallel

Conclusion: **MPI souvent nécessaire**

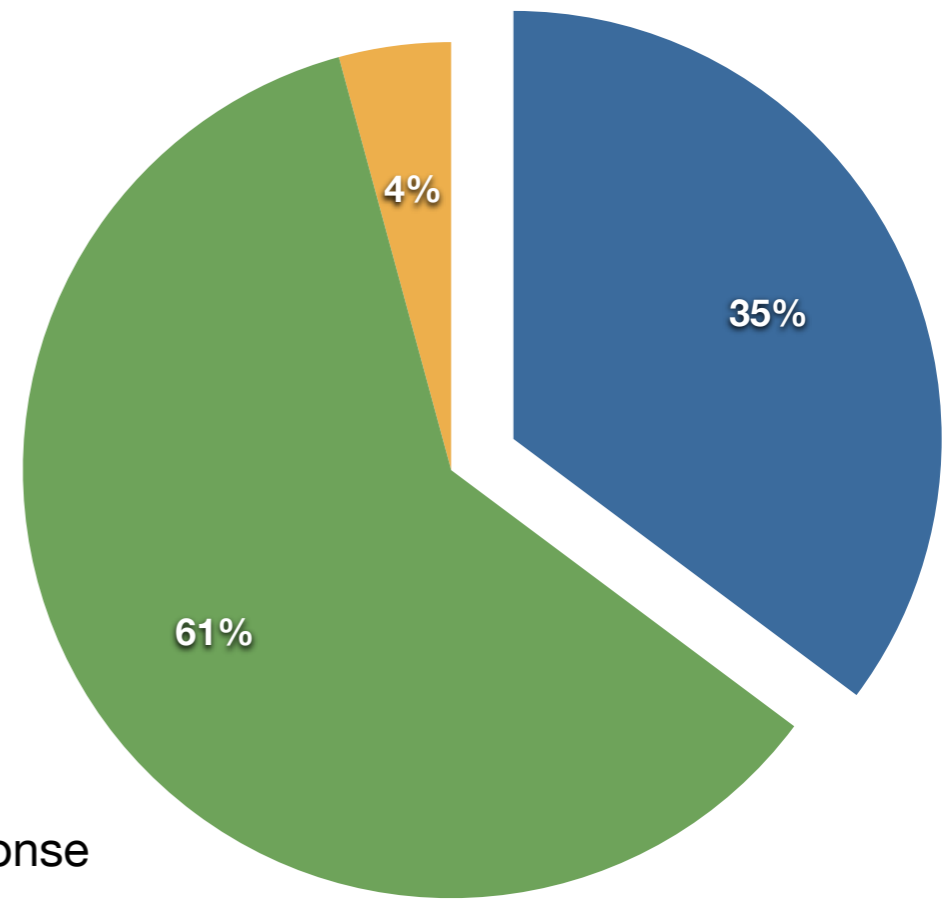
Licences

Vos applications utilisent-elles un **logiciel sous licence** ?

	Nombre	%
Oui	25	35,2
Non	43	60,6
Sans réponse	3	4,2
Total	71	100,0

	Nombre
IDL / ENVI	7
Matlab	5
Geocluster	2
Molpro	2
NAG	1
VU	1
Codior	1
Flac 3D	1
LGPL	1
Compilateur fortran	2
VASP	1

Logiciel sous licence



- Oui
- Non
- Sans réponse

Remarques:

- IDL & Matlab: possibilité de compiler en local et d'utiliser ensuite sur la grille
- Besoin de différencier les cas où il faut le logiciel sur la grille.

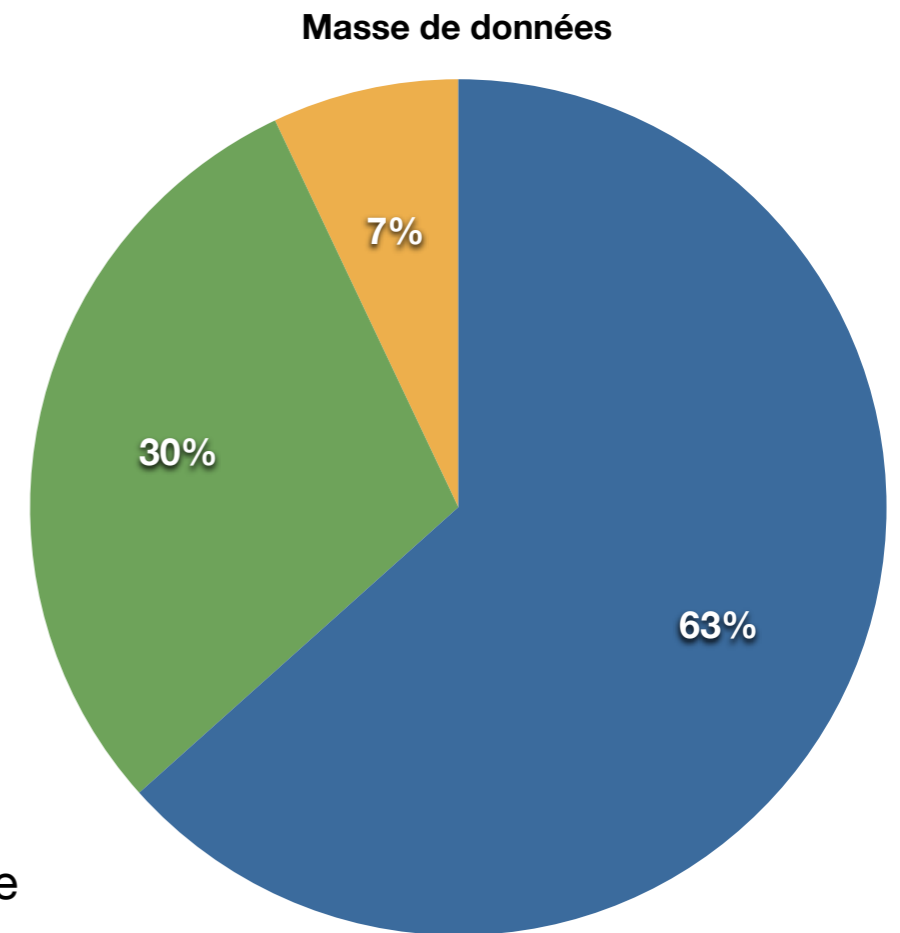
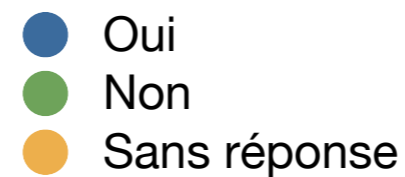
Conclusions

- Besoins d'**informer les utilisateurs sur l'utilisation de ces logiciels sur la grille**
- **Négociations** entre l'institut des grilles et les distributeurs de logiciels

Masses de données

Avez-vous à **utiliser ou produire de grandes masses de données** ou appartenez-vous à des projets impliquant un partage de grandes masses de données avec des collègues distants ?

	Nombre	%
Oui	45	63,4
Non	21	29,6
Sans réponse	5	7,0
Total	71	100,0



Réponses:

- besoins vont de **1 To à 100 To**
- souvent:
 - Archivage dans les centres de données
 - traitement/production sur la grille

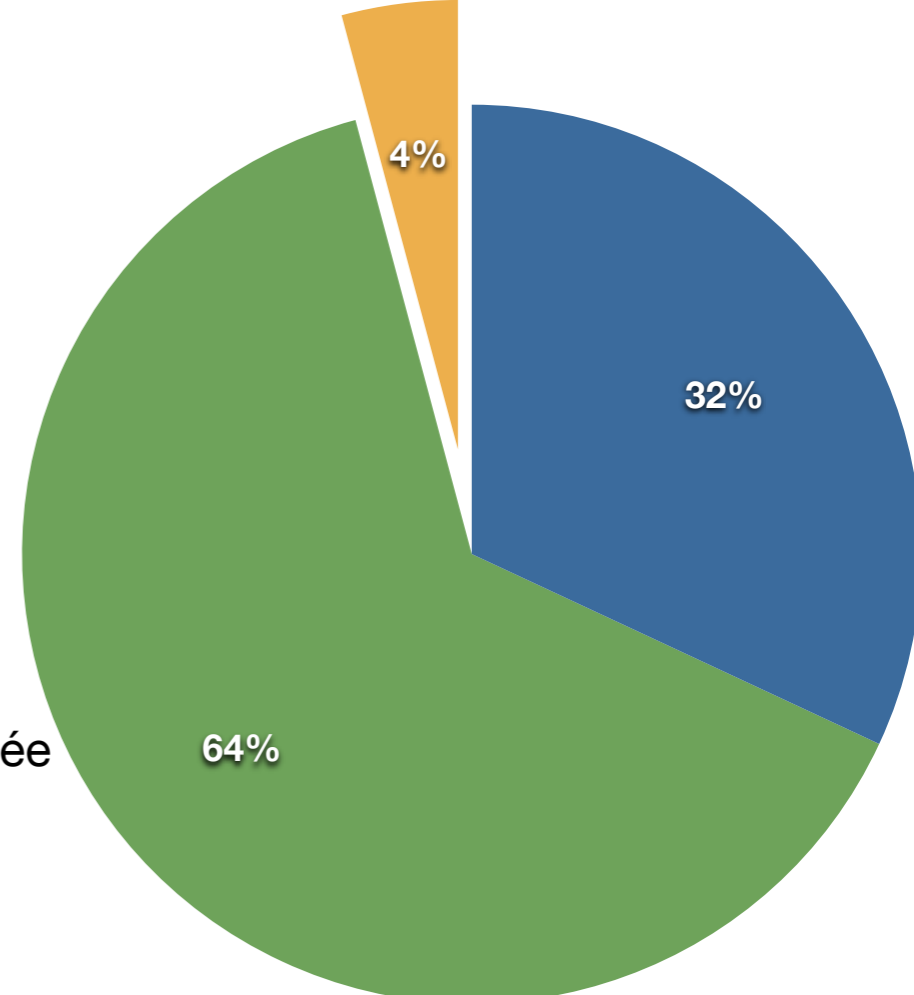
Problèmes de l'accès (délais, intégrité, confidentialité, protocoles) ?

Intérêt des grilles de calcul

Verriez-vous un **intérêt à l'utilisation d'une grille** de production ?

	Nombre	%
Oui et utilisée	23	31,9
Oui mais pas utilisée	46	63,9
Non	3	4,2
Total	72	100

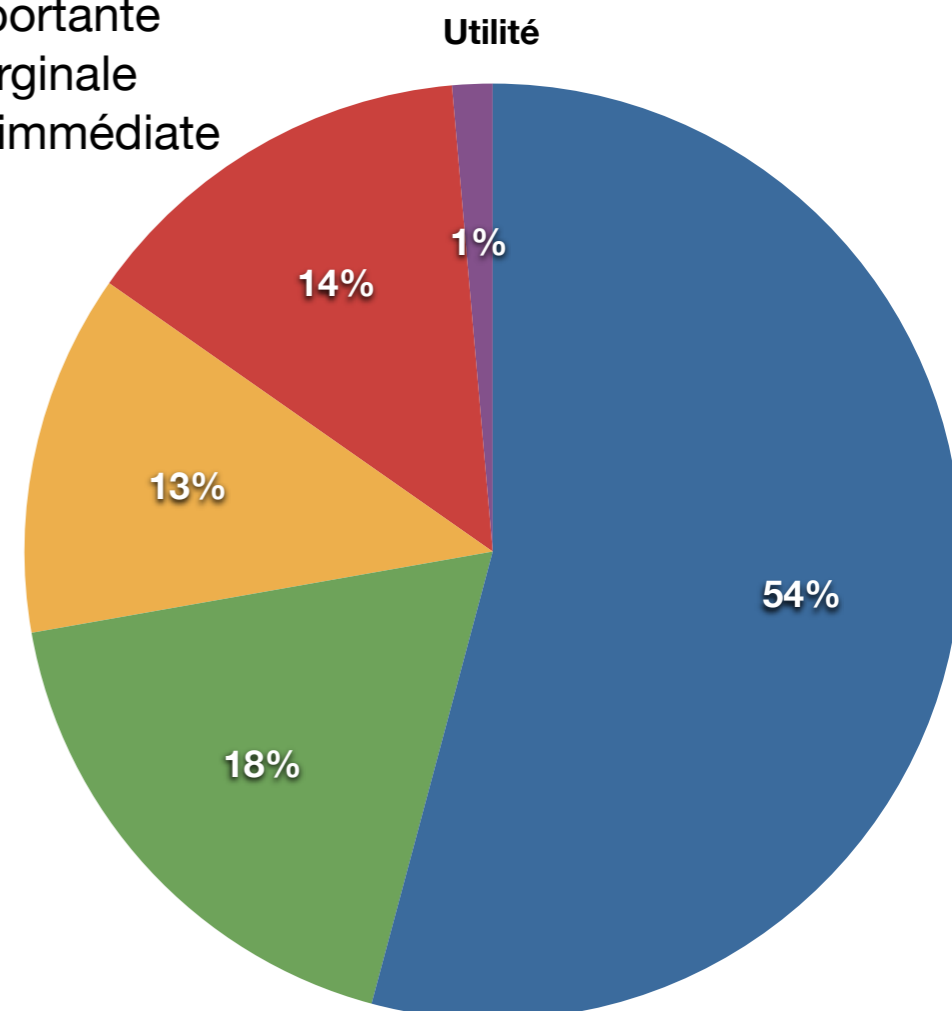
- Oui et utilisée
- Oui mais pas utilisée
- Non



Pensez-vous que cette technologie et les moyens informatiques associés auraient une **utilité** pour vous ?

	Nombre	%
Immédiate et importante	39	54,2
Immédiate et marginale	13	18,1
Indirecte ou non immédiate	9	12,5
Incertaine	10	13,9
Aucune	1	1,4
Total	72	100,0

- Immédiate et importante
- Immédiate et marginale
- Indirecte ou non immédiate
- Incertaine
- Aucune

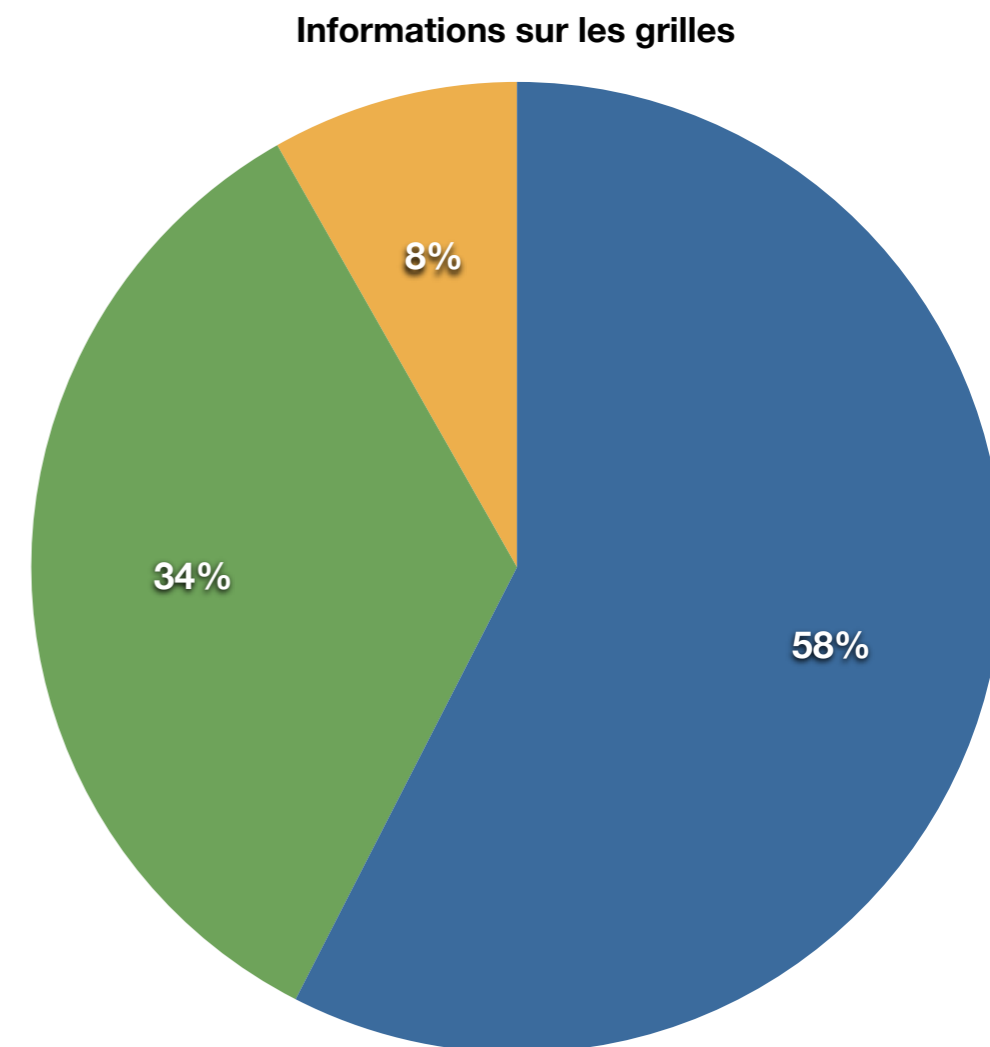


Informations et formations sur les grilles

Pensez-vous être suffisamment **informé sur les grilles de production** et leur apport potentiel à votre recherche ?

	Nombre	%
Un peu	42	57,5
Correctement	25	34,2
Expert	6	8,2
Total	73	100,0

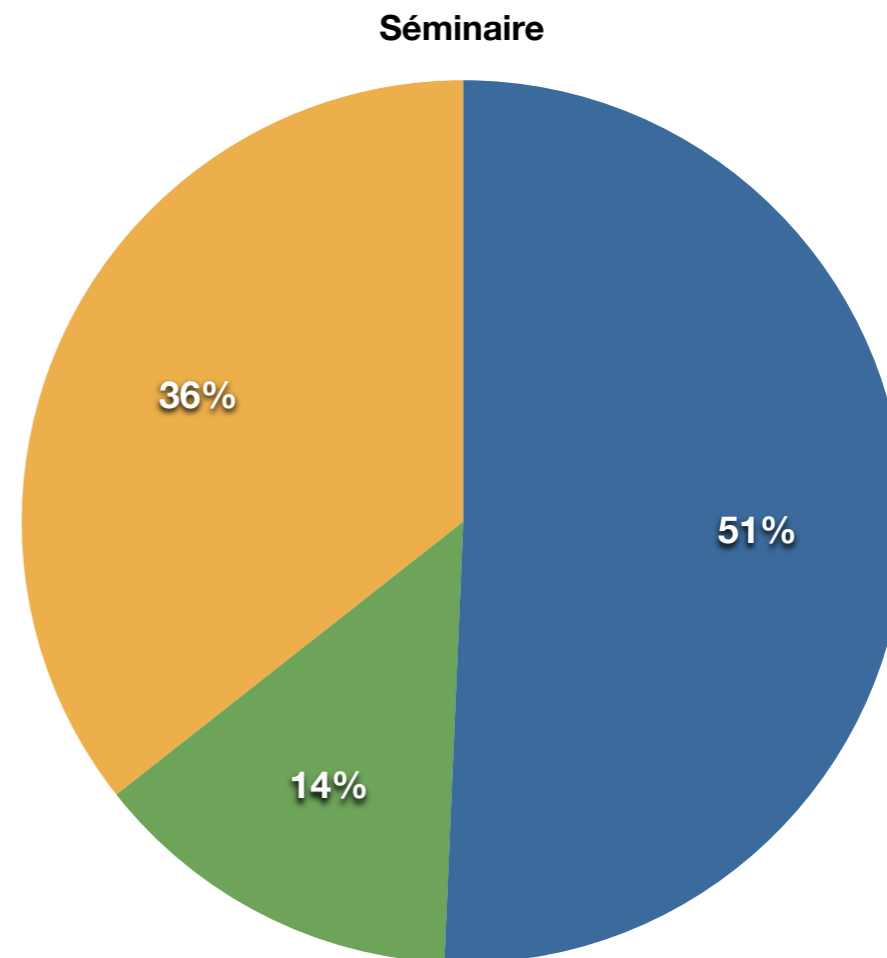
- Un peu
- Correctement
- Expert



Formations sur les grilles

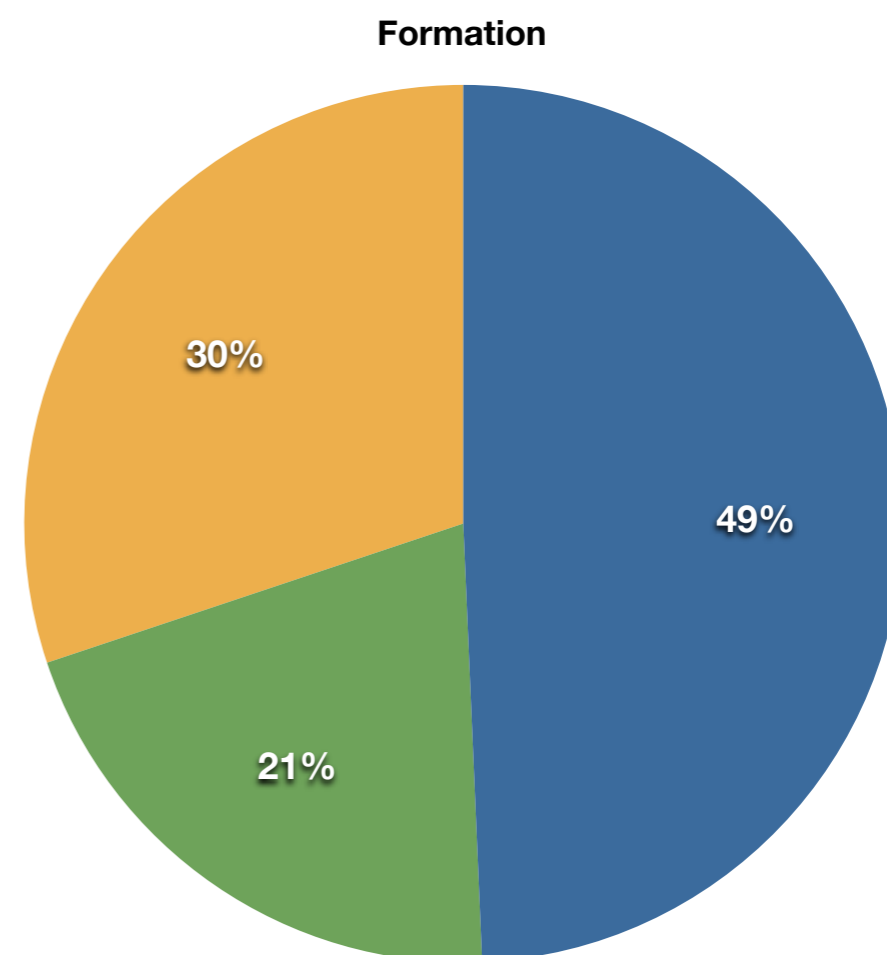
Souhaiteriez-vous assister à un **séminaire d'introduction** aux grilles de production ?

	Nombre	%
Oui	37	50,7
Non	10	13,7
Sans réponse	26	35,6
Total	73	100,0



Etes-vous prêt à suivre une **formation de quelques jours** sur l'utilisation des grilles de production ?

	Nombre	%
Oui	36	49,3
Non	15	20,5
Sans réponse	22	30,1
Total	73	100,0



Conclusion [1]

- **Première large dissémination**
 - Nombre de réponses relativement satisfaisant mais des équipes manquent
 - Réponse par des laboratoires non contactés auparavant
 - Certaines réponses correspondent à une équipe ou groupe
- **Intérêt Scientifique du calcul sur une grille de production**
 - Les réponses au questionnaire montrent qu'une grande variété de disciplines et projets sont intéressés ou se posent des questions d'une possible utilisation de la grille
- **Recensement en calcul sur une grille de production**
 - Résultats qualitatifs et partiels
 - Difficulté de passer à un niveau quantitatif maintenant
- **Points de blocage relevé par le questionnaire pour une utilisation de la grille**
 - Information et formation insuffisantes
 - Moyens humains limités : nombre, temps disponibles
 - Logiciels avec licence
 - Complexité d'utilisation à priori et aussi à posteriori
 - Petit nombre de points de la grille au niveau national d'où difficulté de trouver une aide locale

Conclusion [2]

- Cat 1 : Besoins de calcul facile à satisfaire
- Cat 2 : Thèmes où les groupes sont déjà actifs sur la grille
- Cat 3 : Thèmes où il existe de gros besoins en calcul mais où la grille n'est pas l'outil adapté
- Cat 4 : Thèmes où il existe de gros besoins en calcul où la grille devrait apporter beaucoup et ne le fait pas

Estimation de l'Institut des Grilles :

Cat1: 10%; Cat 2: 15%; Cat 3 : 60% ; Cat 4 : 10%

Nos conclusions après discussion et dépouillement du questionnaire

- **Cat 1** : sous-estimé car beaucoup de calculs pourraient se faire sur la grille mais sont limités par les logiciels
- **Cat 2** : sur-estimé car peu de thématiques impliquées aujourd'hui
- **Cat 3** : sur-estimé mais fort lobbying
 - la prospective super-calculateurs devrait fournir une estimation
- **Cat 4** : rejoint Cat 1 pour les besoins en production et exploration de bases de données, et calculs statistiques

En conclusion il devrait y avoir un fort potentiel d'utilisation de la grille

Conclusion [3]

Projets :

- Actuellement
 - **Nombreux projets individuels** dont certains font partie de projets Européens
 - Applications importantes en sismologie, crue des rivières, évolution des pesticides dans le sol, validation satellitale
- Identifiés et en cours – liste non exhaustive
 - **Traitement des données de grands instruments** (GAIA, Planck, CTA, LOFAR) au sol et embarqués sur satellite
 - Climat : distribution et post-traitement des simulations IPCC/GIEC
 - Nouvelles applications en discussion en hydrologie, geosciences, météorologie...
 - Lien Observatoire Virtuel en Astronomie et Grille

Conclusion [4]

Recommandations et Actions à court et moyen termes

- Continuer la **dissémination du questionnaire**
 - Pour améliorer le nombre de communautés touchées
 - Contacter les personnes qui ont répondu pour affiner les résultats de cette étude
- Utiliser le **matériel de dissémination** développé dans DEGREE et EGEE
- Organiser des **journées d'information et des tutoriaux** par région pour PUE – demande forte
- Utiliser le **support de NA4 dans EGEE III** et des “cluster” A&A et des sciences de la planète
- Avoir des **points de contact par région** si ce n'est localement pour aider les utilisateurs