

# MODELISATION DES PERFORMANCES D'UN SYSTEME DE TELECOMMUNICATIONS ET ANALYSE MASSIVE DE DONNEES

Christine Couloigner

<sup>1</sup> *Christine Couloigner, Nokia, rue Louis de Broglie, 22300 Lannion – christine.couloigner@nokia.com*

**Résumé.** Avec le développement de réseaux de systèmes de télécommunications à grande échelle, il est nécessaire de disposer de modèles basés sur l'exploration de données pour analyser et évaluer et évaluer les performances d'un système en termes de débit et d'utilisation des ressources – tant au niveau des composants individuels que du système dans son ensemble. Le but de cet article est de présenter une méthode de modélisation des performances d'un produit à partir de l'analyse massive de données collectées sur un réseau opérationnel et la construction d'un modèle.

**Mots-clés.** Modélisation, analyse massive de données, exploration de données, performances, télécommunications

**Abstract.** With rapid advances in large-scale networked telecommunications systems data-driven models based on data mining for analyzing and evaluating the performance of a system in terms of throughput and resource utilization -- both at the individual component levels and the system as a whole – are needed. The goal of this paper is to present a method of performance modelling of a product from massive data analysis collected on a live network and building a model.

**Keywords.** Modelling, massive data analysis, data mining, performance, telecommunications.

## 1 La méthode

Cette méthode a été mise au point au fil du temps et des produits. L'article de G.Fiche et F.Le Corre (1990 – référence [1]) puis de G.Fiche, ML Le Gac et F.Thomazeau (1993 – référence [2]) en font une première description. Enfin l'ouvrage de G. Fiche et G.Hébuterne (2003- référence [3]) décrit de manière plus détaillée cette méthode appliquée aux équipements de cœur de réseau.

L'idée générale est de caractériser le système et son environnement par une série de paramètres facilement mesurables afin de déterminer la capacité du système en fonction des caractéristiques de trafic et pour n'importe quel type de configuration du système. Cette démarche est basée sur les trois étapes suivantes :

- 1) Modélisation et évaluation des capacités du système
- 2) Mesures des paramètres du modèle en laboratoire
- 3) Validation du modèle dans un réseau client

## 2 Modélisation et évaluation des capacités des systèmes

La démarche développée ici permet de déterminer la capacité du système en fonction des caractéristiques de trafic et de son environnement de fonctionnement. Elle comporte les trois étapes suivantes :

- 1) Caractérisation de l'environnement en trafic

- 2) Caractérisation des ressources du système
- 3) Evaluation des objectifs de capacité du système

Cette démarche est illustrée par la modélisation des capacités de la station NodeB qui est un des composants du réseau 3G. La figure suivante permet de situer cet élément dans le réseau 3G :

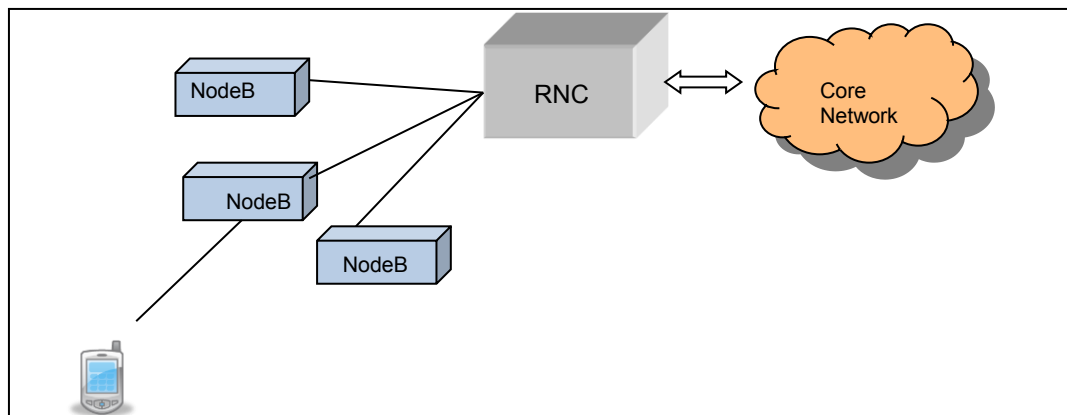


Figure 1-NodeB in 3G network

### 2.1 Caractérisation de l'environnement en trafic

Avec l'arrivée des Smartphones, nous avons observé à la fois une explosion du trafic de signalisation et de données.

L'explosion du trafic de signalisation est due à la multitude d'applications actives sur les Smartphones et déclenchant périodiquement de courtes sessions de récupération de données. Par exemple :

- Application mail : récupération régulière des mails depuis un serveur,
- Application de prévisions météorologiques : mise à jour régulière de la situation météorologique de sa ville préférée,
- Etc. ....

Ceci entraîne également une explosion du trafic de données avec la multiplication terminaux mobiles (Smartphones, tablettes, objets connectés) et de leurs applications. Le trafic de données a ainsi été multiplié par 17 entre 2008 et 2013.

La caractérisation du trafic revient à déterminer les sollicitations en trafic offertes au système comme le trafic de signalisation, de voix ou de données et définir les différents paramètres ayant une influence sur les capacités du système.

Dans le cas du NodeB, les paramètres les plus importants sont :

- Le nombre de messages de signalisation par session,
- La durée d'une session,
- Le débit montant par session,
- Le débit descendant par session.

Il faut d'autre part considérer les différentes fonctions activées ou mode d'exploitation du système comme :

- Type de backhaul (IP, ATM)
- Activation de la fonction Dual Cell,
- Etc ...

### 2.2 La caractérisation des ressources

Elle consiste à :

- Étudier l'architecture du système,
- Définir les lois de service de chaque entité du système. Ces lois seront par exemple des lois de service à temps partagé, service cyclique, service ordinaire).

- Établir la séquence des échanges et des ressources nécessaires par session et pour le type de trafic étudié,
- Lister les ressources nécessaires à l'écoulement de trafic ainsi que les paramètres qui les caractérisent. Ces paramètres sont des données « systèmes » telles que pourcentage d'occupation d'un processeur par message de signalisation,
- Évaluer les taux d'utilisation du système pour type de session,
- Déterminer le taux d'occupation maximum admissible des ressources du système en prenant les temps de réponses et les exigences de qualité de services définies par les clients ou les normes.

A la fin de cette étape, le modèle se définit comme suit dans le cas d'un système comme le NodeB :

*Tableau 1 - caractérisation des ressources*

Paramètres	Ressources				
	R1	.....	Ri	....	Rn
Taux d'arrivée des messages			$\lambda_i$		
Temps de service			$S_i$		$S_i$
Taux d'utilisation du processeur			$\rho_i$		$\rho_i$
Débit en Mb/s					$D_i$

Les ressources sont définies comme un système à file d'attente de type M/M/1 dont les caractéristiques sont les suivantes :

$\rho_i$  : Taux d'utilisation du processeur pour un taux d'arrivée de message  $\lambda_i$ , et définit par la formule suivante :

$$\rho_i = \lambda_i * S_i$$

W : Temps passé dans le système et caractérisé par :  
soit :

w : variable aléatoire décrivant le temps total passé par un message dans le système à file d'attente

$W_q$  : temps passé par un message dans la file d'attente

$W_s$  : temps service d'un message,

$$W = E(w) = W_q + W_s = \frac{W_s}{1-\rho} \text{ et } \sigma_w^2 = W^2$$

La définition des lois de service puis le modèle complet est basé sur la théorie des files d'attente comme décrit dans le livre de Arnold. O Allen (1990) – référence [5].

### 2.3 Evaluation des objectifs de capacité du système

Les paramètres définis précédemment sont dans une première étape basés soit sur des mesures préliminaires réalisées en laboratoire, soit sur des mesures réalisées sur la version logicielle précédente (N-1). Avant de déterminer les objectifs de capacité de la nouvelle version logicielle (N), il est nécessaire d'évaluer l'impact des nouvelles fonctions apportées par celle-ci sur les paramètres du modèle.

Evaluation des impacts des nouvelles fonctions sur les paramètres du modèle

Cette étape consiste à lister les nouvelles fonctions introduites par la version logicielle N et elle est ensuite basée sur deux outils de suivi des impacts :

Un tableau de bord répertorient les nouvelles fonctions et leurs impacts sur les différents paramètres du modèle. Il fait l'objet d'une mise à jour régulière au fur et à mesure de la définition et du développement des fonctions

Un questionnaire partagé avec les personnes en charge de la définition des fonctions et de leur développement. Il répertorie les questions qui permettent à la fois de quantifier les risques d'impact négatifs sur les paramètres du modèle et permet aux experts travaillant sur ces fonctions de minimiser ces risques en les aidants à les identifier.

Application numérique du modèle : calcul des capacités du système pour la nouvelle version logicielle.

Dans cette phase de calcul, on applique les données issues de mesures au modèle défini précédemment. Les taux d'utilisation maximaux sont déterminés afin que les temps de réponse du système et la qualité de service soient conformes aux exigences clients. Le modèle permet de déduire :

- la capacité de chaque ressource
- La capacité de maximale de chaque configuration, en nombre de messages ou débit qu'elle peut traiter pour n'importe quel type d'environnement de trafic.

Tableau 2- Application numérique

Ressource/ Configuration Environnement de trafic	R1	...	Ri	....	Ci
E1					
E2					
.....					
Ei					

Ri : ressource type i

Ci : configuration type i

Ei : environnement de trafic type i

### 3 Collecte des données sur un réseau opérationnel et analyse des données

Les données permettant d'établir le modèle d'évaluation des capacités du système sont ici collectées sur des sites d'un réseau opérationnels.

L'analyse des données consiste à effectuer :

- Une classification des sites et choix d'un site
- Analyse et modélisation du site choisi

#### 3.1 Collecte des données sur un réseau opérationnel

Le protocole de collecte des données est défini de la façon suivante

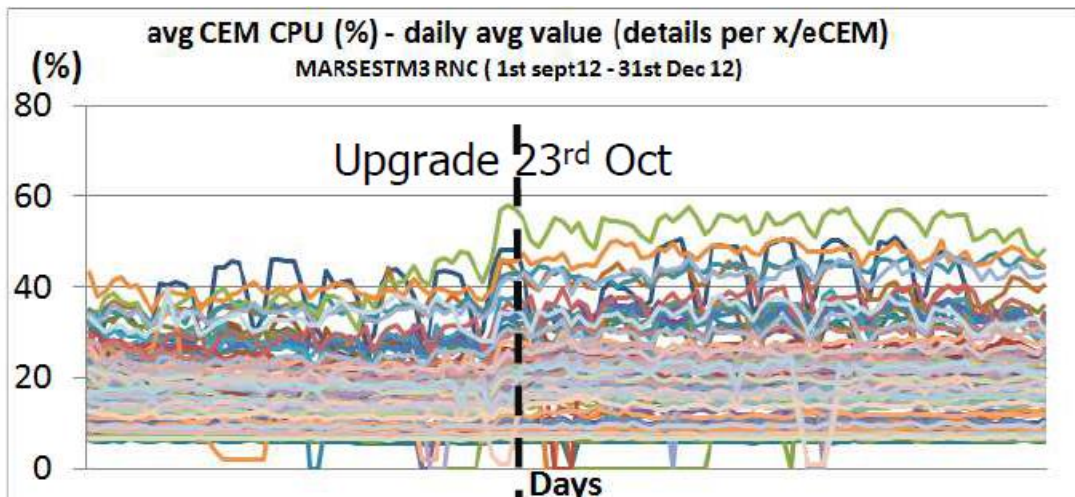
- Collecte des données sur 32 sites opérationnels
- Pour chaque site, une collecte journalière des données est effectuée sur une période de 10 semaines :
  - Pendant 5 semaines avant la mise à jour du système
  - Pendant 5 semaines après la mise à jour du système

### 4 Classification des sites

La classification des sites est effectuée en considérant pour chacun des sites, la variable suivante :

- Taux d'utilisation du processeur ( $\rho_i$ ) appelé ici « Avg CEM CPU- daily average % »
- Débit en Mb/s appelé ici « Average EDCH User Throughput »

Dans un premier temps, nous obtenons les observations suivantes pour la variable taux d'utilisation du processeur( $\rho_i$ ) :



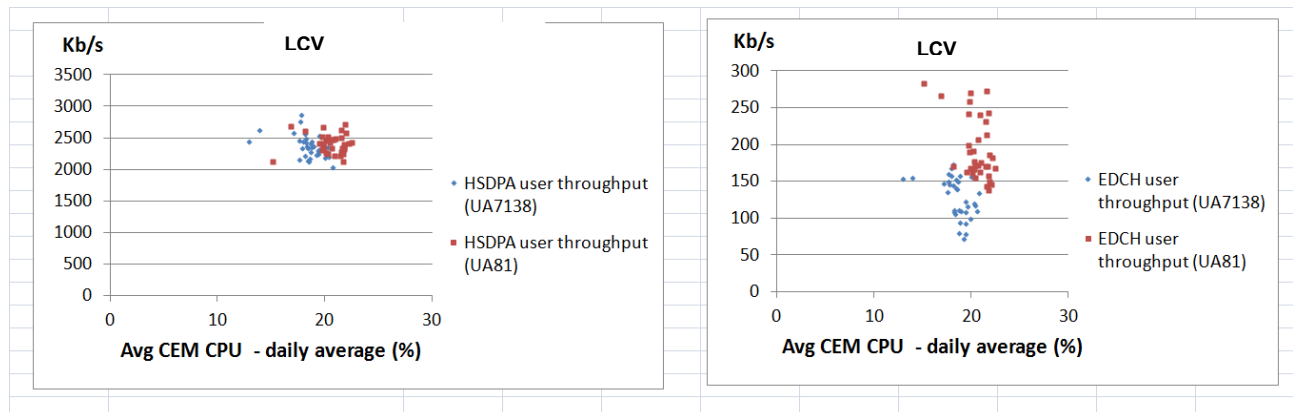
Le graphique ci-dessus permet d'observer que le taux d'utilisation CPU (avg CEM CPU) augmente après la mise à jour :

+ 2.9 pts on avg x/eCEM CPU

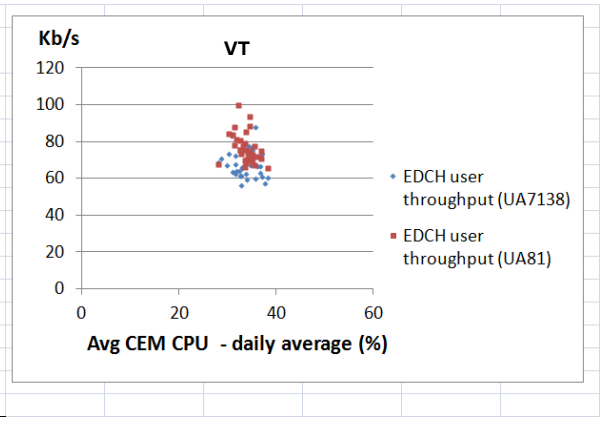
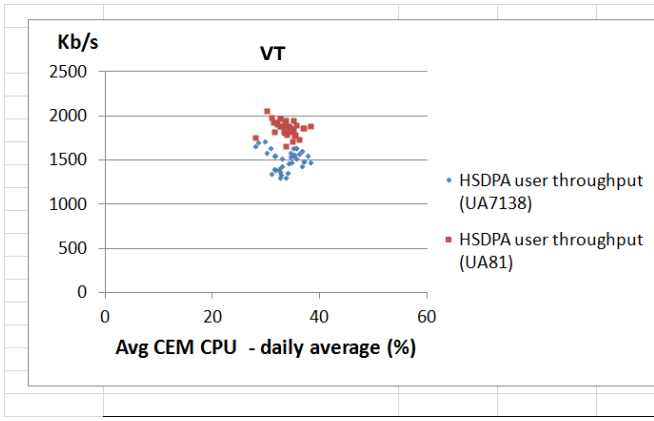
before upgrade	after upgrade
19%	22%

Le but est ici de comprendre la raison de cette augmentation du taux d'utilisation du processeur. On introduit alors les variables de débit :

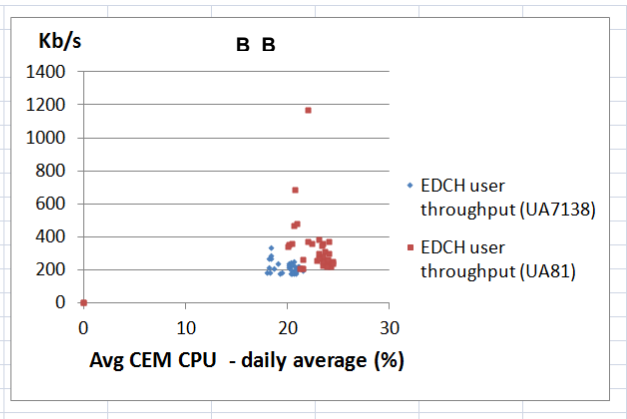
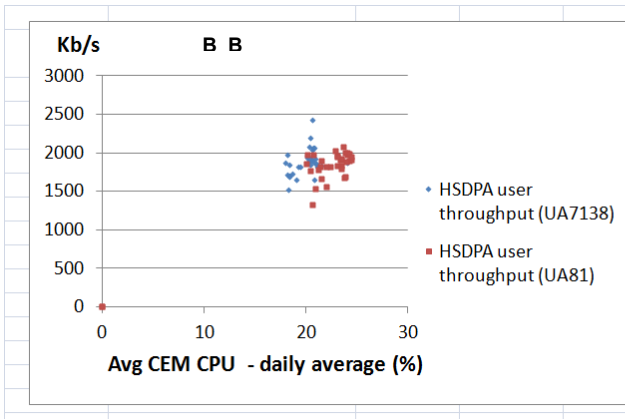
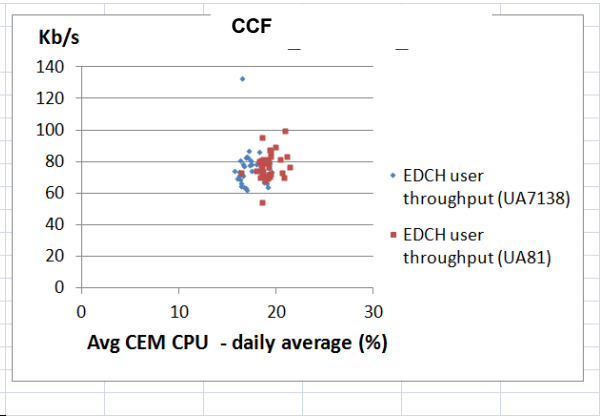
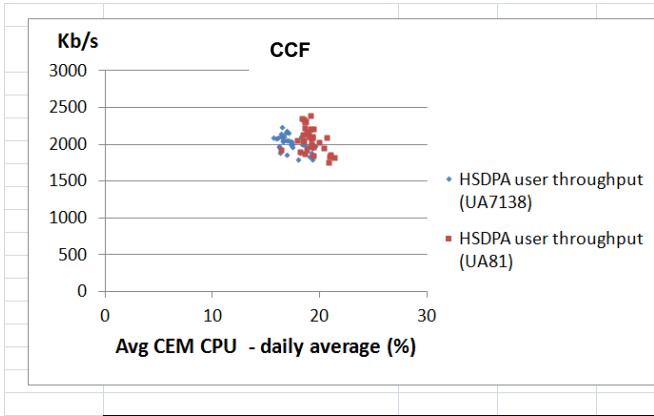
Site 1



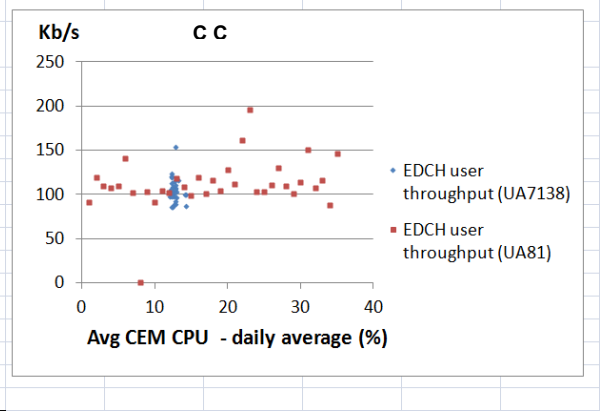
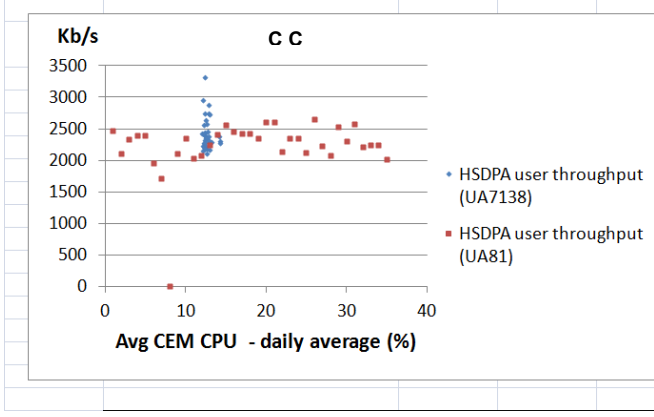
Site 2



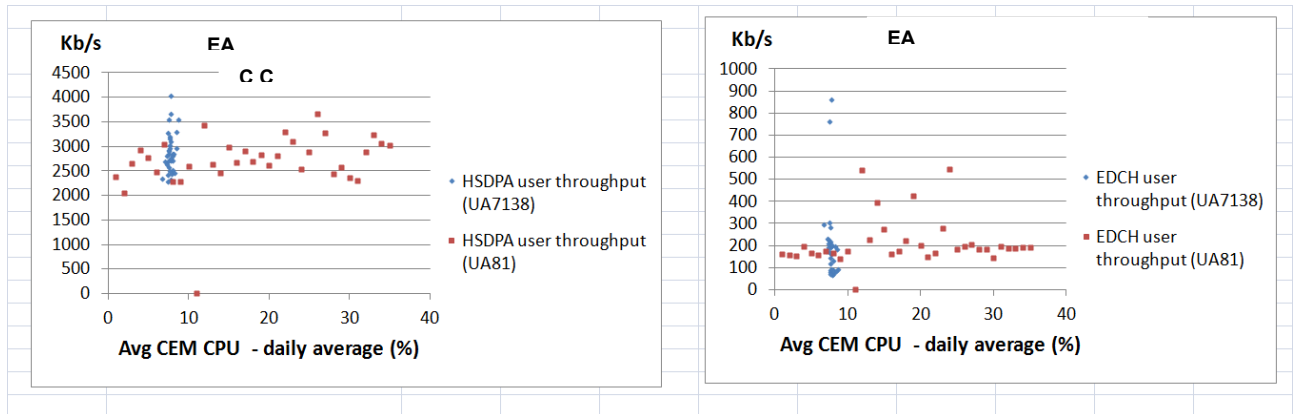
Site 3



Site N



.....



Nous observons des comportements très variées en termes de débit.

« 4.1 classification des sites pour la variable « taux d'utilisation du processeur »

Nous effectuons alors une classification des sites sur la base de la variable

- Taux d'utilisation du processeur ( $\rho_i$ ) appelé ici « Avg CEM CPU- daily average % »

Et nous obtenons le classement suivant :

- Taux d'utilisation du processeur ( $\rho_i$ ) appelé ici « Avg CEM CPU- daily average % »

Site	Modem configuration	Avg CEM CPU Load (PQ3) per modem
BSA	mixity i/xCEM	
B_B	full xCEM	27%
C_C	full xCEM	20%
C_B	mixity i/xCEM	

Selected NodeB :  
Configuration full xCEM and significant CEM CPU load

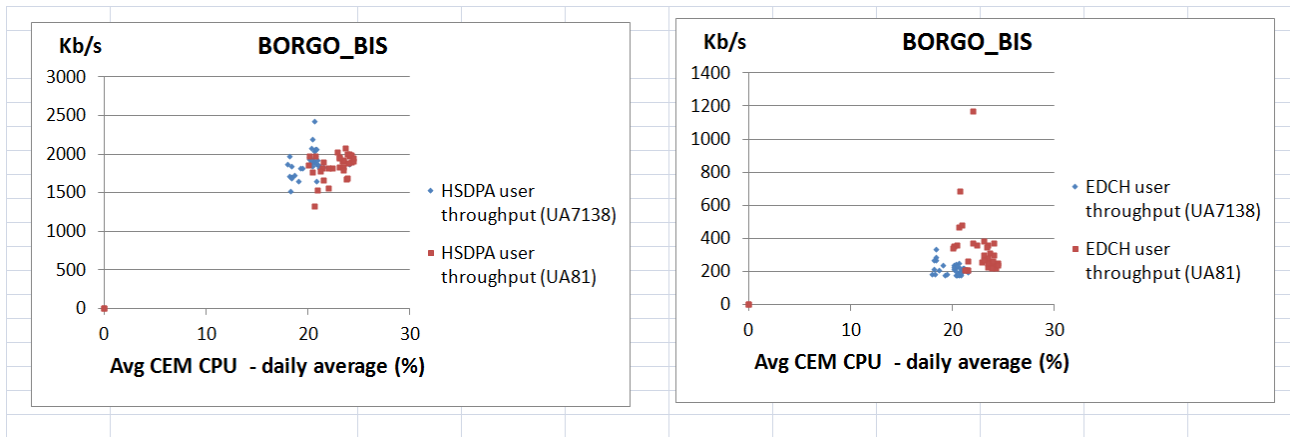
Nous sélectionnons le site B\_B pour une construire la modélisation des performances du système en termes de débit en fonction du taux d'utilisation du processeur.

### 5 Construction du modèle de performance du système

Nous considérons ici deux variables de débit :

- HSDPA user throughput en Kb/s
- EDCH user throughput en Kb/s

Nous obtenons les graphes suivants pour le site B\_B:



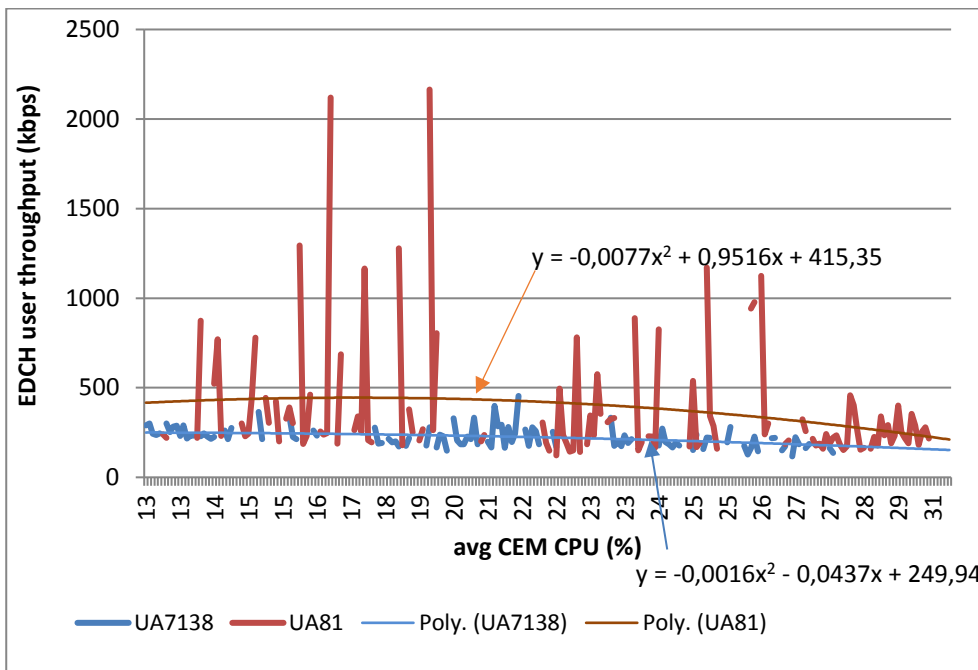
AvgDLThroughputPerUser_HSDPA093_CR(UHSDPA093_CR)(Kbps)	7.1.3.8	8.1.3
Average on the period	1888	1848

AvgULThroughputPerUser_HSUPA009_CR(UHSUPA009_CR)(nb)	7.1.3.8	8.1.3
Average on the period	219	327

Stable

+49%

Nous en déduisons alors le modèle suivant :



Le modèle de performance observé sur le site est donné par les formules suivantes :

- $D_i$  : Débit en Kb/s
- $p_i$  : Taux d'utilisation du processeur en %

1) le modèle avant mise à jour du site

$$D_i = -0,0016 p_i^2 - 0,0437 p_i + 249,94$$

2) le modèle après mise à jour du site

$$D_i = -0,0077 p_i^2 + 0,9516 p_i + 415,35$$



## **6 Conclusion**

Le modèle de performance est construit ici à partir de réseau comportant un faible nombre de sites. Il est intéressant de voir que cette méthode peut s'appliquer facilement à un grand nombre de site et construire un modèle de performance par classification successives de variables.

## **Bibliographie**

- [1] Fiche G. et Le Corre F. (1990), A method for evaluating CCITT performance parameters, application to a switching system, ISS 13, Lannion.
- [2] Fiche G., Sabourin T., Le Gac M.L. et Thomazeau F. (1993), Assessing the performance of the Alcatel 1000 system, Commutation et Transmission n°4, Lannion.
- [3] Fiche G., Hébuterne G. (2003), Traffic et Performance des réseaux de télécoms, édition Hermès Sciences, Lannion- Paris.
- [4] Fiche G., Hébuterne G. (2004), Communicating Systems & Networks : Traffic & Performance, édition Kogan Page Science, Lannion- Paris.
- [5] Allen, Arnold O. (1990), Probability, statistics, and queuing theory with computer science applications, édition Academic press.