ANÁLISIS DE FIABILIDAD DE SISTEMAS APLICANDO TÉCNICAS DE CRECIMIENTO DE FIABILIDAD DEL SOFTWARE

Dña. Amaya Atencia Yépez, D. Luis Redondo López
Métodos y Tecnología de Sistemas y Procesos (MTP)
(lredondo, Amaya.atencia@mtp.es)
www.mtp.es
www.exhaustif.es

Resumen Las técnicas de crecimiento de fiabilidad del software (cuyo acrónimo en inglés es SRGM - Software Reliability Growth Models) se utilizan para estimar el número de fallos latentes, la tasa de fallos y la fiabilidad de un sistema de software. Son técnicas que se han aplicado en el pasado y, por lo general, no se han obtenido los resultados esperados. No obstante, si se aplican conociendo sus restricciones y requisitos, se puede tener excelentes resultados reduciendo considerablemente costes de garantía de productos, lo que implica directamente mejorar la imagen de la compañía y la satisfacción de sus clientes. En este artículo se quiere volver a poner de relieve estas técnicas, indicar los principales problemas que se pueden encontrar y cómo se pueden solucionar. Al ser técnicas estadísticas, sus datos de entrada tienen que reflejar la realidad, de lo contario lo resultados no serán fiables.

Definición de confiabilidad.

Según [MIL-STD 721C] es una medida del grado de operabilidad y capacidad de un sistema para prestar el servicio requerido en cualquier momento de su tiempo de misión, suponiendo su disponibilidad en el instante inicial.

1. INTRODUCCIÓN

Las técnicas de crecimiento de fiabilidad del software (cuyo acrónimo en inglés es SRGM - Software Reliability Growth Models) se utilizan para estimar la tasa de fallos y la fiabilidad de un sistema de software. La necesidad de este tipo de medidas se enmarca dentro del objetivo de cuantificar la confiabilidad de un sistema software, que hace referencia a la "calidad del servicio prestado" de forma que se pueda confiar de una forma justificada en su servicio.

La confiabilidad (dependability) no se mide directamente, sino a través de sus atributos, los cuales, expresados desde una perspectiva RAMS son: fiabilidad (reliability), disponibilidad (availability), mantenibilidad (maintainability) y seguridad (safety).

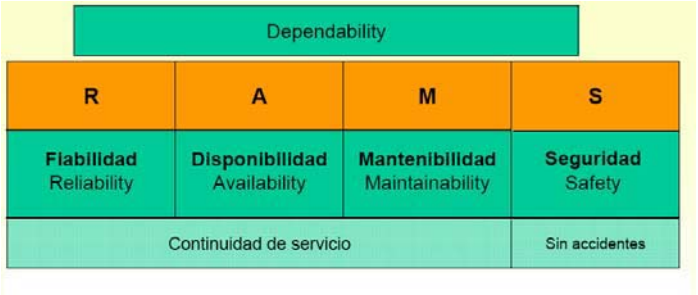

Figura 1

Una medida de confiabilidad importante es la tasa de fallos (i.e. el número de fallos por unidad de tiempo) que sirve para evaluar la frecuencia de fallos de un sistema tal y como es percibida por el usuario. En hardware, la fluctuación normal de la tasa de fallos suele ser la conocida curva de la bañera que se refleja en el gráfico siguiente:

Sin embargo, el comportamiento de la tasa de fallos para sistemas de software es diferente, ya que cuando el sistema se encuentra disponible para el usuario se puede considerar que la tasa de fallos permanece constante en el tiempo. Ahora bien, durante el periodo de desarrollo de un sistema de software este comportamiento no es así, ya que la tasa fallos no es constante en el tiempo. De hecho, durante la fase de pruebas la tasa de fallos debería ser decreciente y aproximadamente constante durante su fase de producción (ver gráfico).

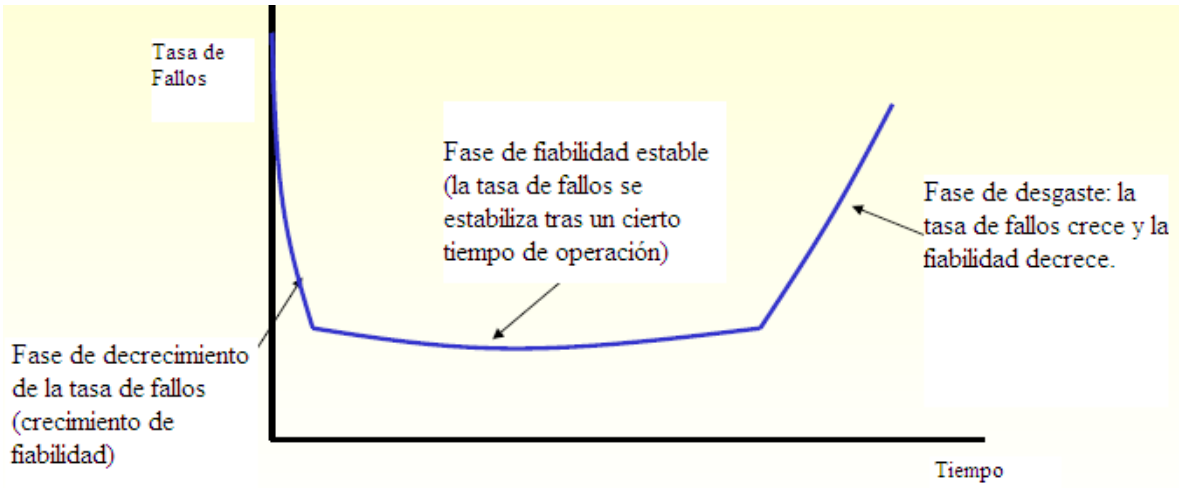

Figura 2. "Curva de la bañera"

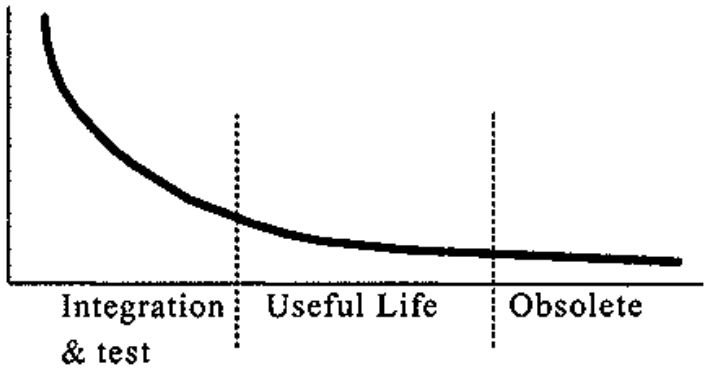

Figura 3. "Tasa de fallos para sistemas software"

Idealmente, durante la fase de pruebas, el software experimenta las siguientes variaciones en la tasa de fallos:

A medida que la fase evoluciona, el equipo de pruebas adquiere una mayor experiencia en el funcionamiento del sistema y es capaz de ir detectando un número mayor de defectos.

Hacia el final de las pruebas se espera que sólo aquellos defectos que son más difíciles de detectar permanezcan en el software.

Según este modelo la tasa de fallos tendría un comportamiento similar al mostrado en la gráfica siguiente (en el eje de las abscisas se registra el tiempo acumulado de pruebas):

Figura 4. "Comportamiento de la tasa de fallos durante la fase de pruebas"

2. TÉCNICAS SRGM

Aplicando técnicas de SRGM durante la fase de pruebas del desarrollo de un sistema de software, se podrá predecir el número de fallos que tendrá un sistema en la fase de producción, de forma que el gestor del proyecto dispondrá de una útil herramienta de decisión para:

1. Establecer si el conjunto de pruebas al que ha sido sometido el sistema ha resultado suficiente, de forma que se pueda autorizar su paso a producción.
2. Reducir la posibilidad de que ocurra una incidencia grave en producción.
3. Estimar la tasa de fallos del sistema.
4. Estimar el momento, durante la fase de pruebas, en que se alcanzan los objetivos de fiabilidad del sistema. Por ejemplo, se puede haber determinado en los requisitos del sistema la calidad del software como: "Después de las pruebas, con un nivel de confianza del 95% deben quedar menos de 10 errores residuales en el sistema". Utilizando estas técnicas se podría estimar cuándo se alcanzará este objetivo.

Modelos de crecimiento de fiabilidad

Los modelos de crecimiento de fiabilidad se clasifican según el tipo de variable aleatoria bajo estudio: TBF (Time Between Failures - Tiempo entre Fallos) o FC (Failure Count – Número de fallos observados por unidad de tiempo).

Modelos cuya variable estudiada es el TBF	*Modelos cuya variable estudiada es el FC*
Los datos observados de estos modelos son los valores de tiempo entre fallos. El parámetro estimado es el tiempo medio hasta el próximo fallo. Están basados en distribuciones exponenciales. Estos modelos establecen que, a medida que se vayan eliminando los defectos del modelo el TBF se irá incrementado. Algunos ejemplos: Jelinski-Moranda o Musa.	Los datos observados de estos modelos son nº de fallos / unidad de tiempo. El parámetro estimado es el número de fallos en un intervalo específico de tiempo cuya amplitud se fija a priori. Están basados en distribuciones NHPP. Estos modelos suponen que a medida que los defectos sean detectados y eliminados, el número de fallos/unidad de tiempo decrecerá. Algunos ejemplos: Musa-Okumoto, Logarithmic o Scheneidewind.

La calidad de las predicciones del modelo de fiabilidad de un sistema se ve afectada por dos tipos de incertidumbres sobre el comportamiento de los fallos detectados:

Primero: no se conoce el efecto que sobre la fiabilidad tendrá la reparación de un error (de forma que se pueden introducir nuevos errores en el software), o
Segundo: Se desconoce el efecto del entorno operacional, esto es, no sabe cuando ni qué inducirá a la detección de un fallo.

Existen más de un centenar de modelos, pero muchos de ellos no han podido ser probados en entornos operacionales con datos reales. Los modelos FC han resultado muy exitosos a la hora de modelar muchos sistemas reales. Y, además, abordan ambas incertidumbres de una forma más realista que los modelos TBF al contemplar cuestiones como por ejemplo:

Los errores que se pueden quitar e introducir en el software cuando se repara un fallo,
El conocimiento que sobre el sistema van adquiriendo los ingenieros a medida que se avanza en las pruebas o
Que se encuentran fallos más sencillos de detectar en las etapas iniciales del proceso de pruebas.

Los modelos de crecimiento de fiabilidad son modelos paramétricos que están definidos por el valor de ciertos parámetros, cada uno de los cuales tiene un significado concreto. Por ejemplo, el número esperado de fallos para el modelo de Scheneidewind viene expresado por la siguiente función:

$$m(i) = \frac{a}{b}(1 - \exp(-bi)) \quad \text{(a)}$$

Los parámetros que definen este modelo son: a, número de fallos al comienzo de las pruebas y b, tasa de fallos por unidad de tiempo.

Estos parámetros se pueden estimar dinámicamente mediante métodos de inferencia estadística, como por ejemplo por Máxima Verosimilitud o estimación Bayesiana de más reciente aplicación, los cuales se nutren de los datos reales recogidos a lo largo de la fase de prueba.

Para el conjunto de datos observados existe la posibilidad de utilizar diferentes modelos para realizar las estimaciones. En el siguien-

te gráfico (FC, nº intervalo/fallos) se comparan los datos observados (puntos verdes) con los predichos por varios modelos una vez estimados sus parámetros.

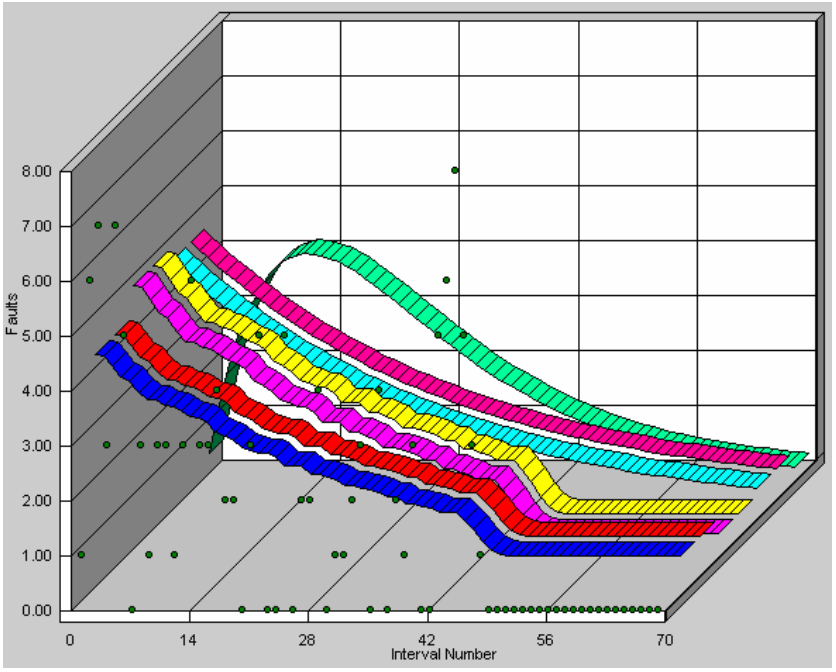

Figura 5. "Comparación de los datos observados con los estimados por los modelos"

Sin embargo, no se puede establecer a priori, y sólo basándose en las hipótesis que lo definen y en las características del entorno operacional, el modelo que mejor se va a ajustar a los datos observados y que va a producir las estimaciones más precisas. De entre todos los posibles modelos es necesario elegir aquel que mejor "modela" el sistema real bajo prueba. Para este fin se utilizan una serie de indicadores que comparan algunas características de los modelos:

1. Bondad de ajuste de las predicciones con las observaciones.
2. Capacidad predictiva del modelo a corto o largo plazo.
3. Tendencia al optimismo o al pesimismo de las predicciones.
4. Precuential likelihood.

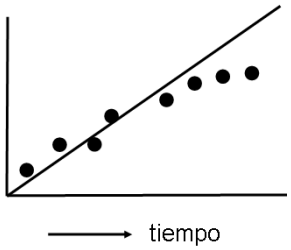

Figura 6.

A modo de ejemplo en este gráfico u-plot, se puede apreciar como las estimaciones de un modelo se ajustan bastante bien a los datos observados en las etapas iniciales de la fase de pruebas, pero no es así al final.

Si ninguno de los modelos resulta suficientemente bueno es posible mejorarlos mediante técnicas de recalibrado, aunque también se puede probar a eliminar el ruido directamente de los datos observados utilizando técnicas de "smoothing". Resulta muy importante también conocer las hipótesis que

definen cada uno de los modelos, ya que, según éstas, algunos son más apropiados dependiendo del momento de la fase de pruebas o del entorno operacional.

Existen herramientas automáticas que pueden servir como soporte a estos análisis, como por ejemplo: SMERFS o CASRE.

Finalmente, es importarte señalar que los datos recogidos están íntimamente vinculados a la estrategia de pruebas que se utilice. Existen diferentes métodos para generar casos de prueba y distintas aproximaciones para detectar los fallos como por ejemplo el uso de perfiles operacionales los cuales ofrecen una aproximación estadística con el objetivo de simular el uso más típico del sistema. El uso de una u otra estrategia de prueba provocará alteraciones en la frecuencia y severidad de los fallos observados, lo que indica que los modelos elegidos deberán ser también adecuados a esta estrategia.

3. APLICACIÓN PRÁCTICA

A continuación, se describen las dificultades y los resultados obtenidos en la aplicación práctica de técnicas de Crecimiento de Fiabilidad SW para el desarrollo de sistemas en una gran compañía de desarrollo de software en España. El entorno de trabajo fue la fase de pruebas de integración de muy diversos proyectos de sistemas, previamente a su paso a producción. La dificultad de la aplicación de estas técnicas residió fundamentalmente en los siguientes tres factores:

- El gran dinamismo de las actividades comerciales de la compañía. Esto implicaba que de forma muy frecuente y rápida se lanzaran al mercado nuevos productos o ampliaciones de los ya existentes, de forma que los proyectos bajo prueba no solían ser demasiado grandes y, por tanto, los tiempos planificados para llevar a cabo las pruebas eran extremadamente ajustados y los equipos de prueba solían trabajar en varios proyectos de forma simultánea. La consecuencia directa sobre el análisis de fiabilidad era el riesgo de disponer de pocos datos para realizar el estudio.

- La complejidad de la estructura de los proyectos. Normalmente eran varios los equipos involucrados en las pruebas. De hecho, teóricamente había una metodología de pruebas común, pero en la práctica cada equipo era dirigido de manera independiente por un jefe equipo con sus propios conocimientos, experiencias y forma de trabajo. Además, cuanto mayor era el número de sistemas implicados en el proyecto, más difícil resultaba conocer por ejemplo la cobertura funcional de las pruebas, el grado de impacto del fallo de un sistema en el resto, el tiempo de ejecución de un caso de prueba o la fiabilidad global.

- La dificultad para recoger los datos de prueba relacionados con el tiempo de uso del sistema. Debemos tener considerada una métrica fiable de esfuerzo real de prueba, esto es, soportada por una herramienta software. Este hecho implicaba un alto riesgo de imprecisión en los resultados del análisis, ya que las dos principales claves del éxito en el uso de las técnicas de crecimiento de fiabilidad son por una parte, realizar las pruebas en un entorno lo más similar posible al de producción, contando con un diseño de casos de prueba que ejerciten el sistema en la forma más probable a ser usada por el usuario y por otra parte, recoger los tiempos exactos de ejecución del sistema.

Mediante la definición e implantación de procedimientos, métricas y herramientas adecuadas y, sobre todo, gracias a la implicación y colaboración de todo el departamento de pruebas de integración, se pudie-

ron superar los riesgos de los dos últimos puntos. Sin embargo, las dificultades se siguieron encontrando en la escasez de datos proporcionados por los proyectos con un tiempo de pruebas corto. De hecho, por regla general, los proyectos que no estuvieran, al menos, dos semanas bajo prueba, no tenían posibilidad de estimación alguna y, normalmente aquellos que no estuvieran al menos cuatro semanas no aseguraban una bondad de ajuste con un nivel de confianza suficientemente alto. Una de las posibles soluciones a este problema sería utilizar técnicas de estadística Bayesiana para la estimación de los parámetros de los modelos, ya que permiten obtener resultados con un conjunto de datos menor que el utilizado por las técnicas convencionales de inferencia estadística. A la hora de ejecutar los modelos, los datos de entrada que se manejaron fueron los siguientes: tiempo de máquina invertido en la ejecución de los casos de prueba y las incidencias registradas (sistema afectado, severidad y momento en que se produjo). Los datos fueron recogidos de la forma más automatizada posible, minimizando en lo posible el factor humano, que es una fuente muy probable de introducción de imprecisiones. Tras su recogida fueron sometidos a un proceso de depuración, filtrado, agrupación y normalización. Al término de este proceso se disponía de un conjunto de intervalos de duración conocida y con el número de fallos adecuadamente distribuido en cada uno. Con esta información sólo se pudieron ejecutar modelos cuya variable estudiada era el FC (esto es contabilizarán el número de fallos por intervalo), ya que fueron descartados por imprecisos los datos observados que aportaban el tiempo entre las ocurrencias de los fallos (TBF).

En resumen, aplicando técnicas de crecimiento de fiabilidad a aquellos proyectos con un tiempo de pruebas suficientemente alto, la compañía pudo beneficiarse de la siguiente información:

- Estimación del número de fallos residuales una finalizada la fase de pruebas de integración (lo cual es una herramienta muy potente para conocer si el paso a producción se va a realizar en el momento adecuado).
- Estimación de la tasa de fallos.
- Estimación de la fiabilidad del producto.
- Conocimiento del error estándar en la estimación.
- Conocimiento de la criticidad (o severidad) esperada de los fallos residuales.
- Estimación del tiempo necesario añadido de pruebas para reducir el número de fallos residuales.
- Conocimiento de la distribución del esfuerzo de pruebas (tanto en tiempo como en nº de casos ejecutados) a lo largo de la fase de pruebas.
- Conocimiento de la distribución de los fallos encontrados a lo largo de la fase de pruebas (lo que puede servir como instrumento para evaluar la calidad del plan de pruebas).

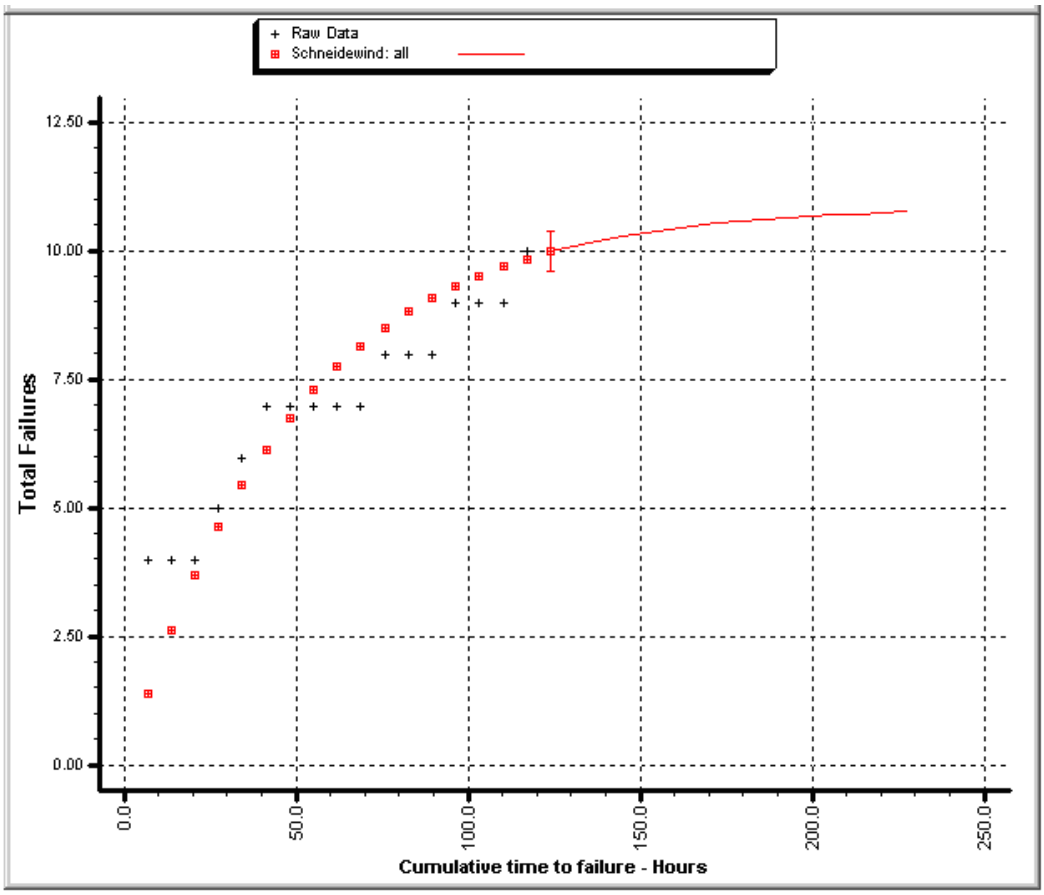

Gráfico número acumulado de fallos por tiempo acumulado con una predicción para 15 intervalos de 6.89h/intervalo (representado en el modelo por la gráfica continua).

La pendiente de la curva que representa el número total de fallos encontrados es bastante pronunciada hasta casi el final de los 18 intervalos de prueba. Una vez alcanzado este punto y según el modelo, durante un periodo de tiempo similar a un tercio del periodo de pruebas de integración el ritmo de crecimiento irá decreciendo hasta hacerse nulo cuando ya se hayan encontrado todos los fallos esperados

Figura 7. Gráfico del Tiempo Acumulado hasta fallo (generado por la herramienta CASRE)

4. CONCLUSIONES

Las técnicas de crecimiento de fiabilidad del software, se utilizan para estimar la tasa de fallos y la fiabilidad de un sistema de software durante la fase de pruebas de su desarrollo. Además, se podrá predecir el número de fallos que tendrá un sistema en la fase de producción, de forma que el gestor del proyecto dispondrá de una útil herramienta de decisión para:

1. Establecer si el conjunto de pruebas al que ha sido sometido el sistema ha resultado suficiente de forma que se pueda autorizar su paso a producción.
2. Reducir la posibilidad de que ocurra una incidencia grave en producción.
3. Estimar el momento, durante la fase de pruebas, en que se alcanzan los objetivos de fiabilidad del sistema.

MTP aplicó técnicas de crecimiento de fiabilidad en muy diversos proyectos en una compañía de desarrollo de SW durante la fase de pruebas de integración de sistemas, previamente a la puesta en producción. En algunos proyectos, debido al escaso tiempo de pruebas, no se pudieron obtener estimaciones precisas. Sin embargo, en aquellos proyectos con un tiempo de pruebas suficientemente alto, pudo obtenerse una información muy útil para la gestión de la fiabilidad de sistemas tanto en la fase de pruebas como en producción.

5. AGRADECIMIENTOS

Agradecemos al Departamento de Pruebas de Integración de Orange que nos haya permitido publicar los resultados más destacables tras aplicar estas técnicas de SRGM en sus instalaciones.

6. REFERENCIAS

[1] J. D. Musa "Software Reliability Engineering". Second Edition. 2004

[2] C. Kaner, J. Falk, H.Q. Nquyen, Testing Computer software (2nd Ed), International Thomson Computer Press, 1993.

[3] M.A. Friedman, P.Y. Tran y P.L. Goddard, "Reliability of Software Intensive Systems", Noves Data Corporation, ISBN: 0-8155-1361-5, 1995.

[4] R. S. Presuman, "Ingeniería del Software: Un enfoque práctico". 5ª Edición. Mc Graw Hill.

[5] J. Pukite "Modeling for Reliability Analysis". Wiley-IEEE
PRess, 1998

[6] Munson, J; Khoshgoftaar, T. Handbook of Software Reliability Engineering". Wiley. 1996.

[7] Carreira, J; Costa, D. Dependability Validation, Evaluation and Testing of Safety-Critical Aerospace Systems. DASIA'99. May 1999.