

## システム基盤構築工数見積もりモデルの継続的改善と普及展開

井ノ口 伸人<sup>†1</sup>  
渡辺 絢子<sup>†1</sup>  
渡辺 真太郎<sup>†1</sup>

大杉 直樹<sup>†1</sup>  
吉野 順<sup>†1</sup>  
戸村 元久<sup>†1</sup>

伏田 享平<sup>†1</sup>  
藤貫 美佐<sup>†1</sup>  
木谷 強<sup>†1</sup>

本稿では、アプリケーションソフトウェアを稼働させるための土台であるシステム基盤構築の工数見積もりモデルの継続的な改善と普及展開について述べる。アプリケーションソフトウェア開発の工数見積もりモデルについては多くの研究で有効性が報告されているが、企業のITシステム開発実務で広く利用されているとは言い難い。システム基盤構築については工数見積もりモデルの研究は少なく、利用事例も多くない。本稿では、NTTデータにおけるシステム基盤構築工数見積もりモデルの継続的な改善と利用範囲の拡大について事例を報告する。規模を表すパラメータ8種類、基盤構築の難易度や能力を表すパラメータ12種類を、プロジェクトマネージャからアンケート調査でデータを収集した。重回帰分析で予測モデルを作成し、社内、国内グループ会社へ普及展開を行った。2008年度2つの部署のプロジェクト10件のデータから作成した見積もりモデルを文書化して当該部署に提供していた取り組みは、2012年度20の部署から収集したデータ43件から作成した予測モデルをWebアプリケーションに組み込んで全国グループ会社へ提供するに至った。2008年度に0.31であったモデルの相対誤差中央値は、2012年度に多少悪化した。一方、見積もりモデル作成に利用したデータに基づき、2008年度はサーバ台数が3~15台の小規模システムのみを適用可能プロジェクトとしていたが、2012年度にサーバ台数3~70台の中大規模システムまで適用可能範囲を拡大した。2008年度5名であったモデルの年間利用者数は、2012年度には479名まで増加した。

## Continuous Improvement and Deployment of A Series of Effort Estimation Models for System Platform Development

Nobuto Inoguchi<sup>†1</sup>  
Junko Watanabe<sup>†1</sup>  
Shintaro Watanabe<sup>†1</sup>

Naoki Ohsugi<sup>†1</sup>  
Jun Yoshino<sup>†1</sup>  
Motohisa Tomura<sup>†1</sup>

Kyohei Fushida<sup>†1</sup>  
Misa Fujinuki<sup>†1</sup>  
Tsuayoshi Kitani<sup>†1</sup>

This paper reports continuous improvement and deployment of a series of effort estimation models for system platform development. Many studies have reported the effectiveness of effort estimation models; however few of them have used in companies' practical system development. This paper reports an example of deployment of a series of estimation models in NTT DATA Corporation with continuous improvement. In order to derive the model, we first collected 8 metrics for sizing and 12 metrics for measuring difficulty and team capability by a questionnaire for 10 project managers. We then derived the statistical model by regression analysis with the collected data. Although median of relative error of the model was 0.31 in 2008, it had been a little degraded to 0.34 in 2012; on the other hand, its applicability had been largely expanded from small-scale system development consisting of 3 to 15 servers, to medium- or large-scale system development consisting of 3 to 70 servers. While we first only provided the documented model to 5 practitioners in 1 company section, now we had developed a web application with the estimation model. This web application was used by 5 users only in 2008; but it had been grown up to 479 users in 2012.

### 1. はじめに

ITシステムの開発に要する工数を見積もるため、多くの研究が行われてきた1)4)6)9)15)28)30)39)。特に開発するシステムの規模や工数に影響を与える要因をデータとして収集し、重回帰分析などで工数を算出する予測モデル(工数見積もりモデル)は、過度の楽観視による大幅な見積もり誤りを防止できるなどの有効性が報告されている3)8)12)17)18)22)38)。また、開発担当者が経験のないツールやフレームワークを利用する場合など、不確実性が大きい状況では、専門家が行うよりも見積もりの精度が高い(正確に見積もれる)ことが報告されている5)13)26)29)35)。

一方、企業のシステム開発実務で工数見積もりモデルが広く利用されているとは言い難い2)11)21)23)24)40)。著者らが所属するNTTデータでも、FP法1)を始めとする工数見積もり支援

専門部隊が全社研究開発部門に存在していたものの、収集したデータの基本統計量をまとめた報告書を年次発行するに留まっていた。大幅な見積もり誤りによる赤字プロジェクト発生の度に工数見積もりの重要性は議論されるものの、本稿の取り組み以前に工数見積もりモデルの実務利用を全社的に進めた事例はなかった。

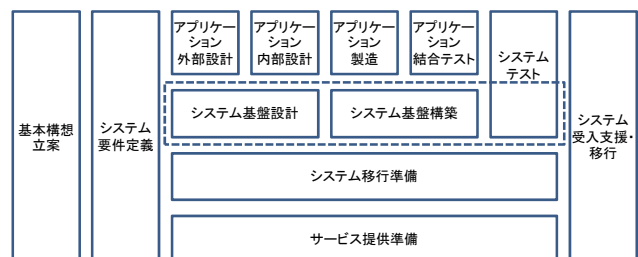


図1 NTTデータのプロセス定義におけるシステム基盤構築  
Figure 1 Definition of System Platform Configuration in NTT DATA Corporation's Development Processes

<sup>†1</sup> 株式会社NTTデータ 技術開発本部  
Research and Development Headquarters, NTT DATA Corporation

表 1 日本のシステム開発企業における工数見積もりモデルの利用事例  
 Table 1 Practices of Effort Estimation Model in Japanese System Development Companies.

企業名	見積対象	見積もりモデルの概要	利用実績
富士通	Web システムが対象, 開発全体工数	Function Scale (簡易 FP) を規模パラメータとする COCOMO II 4) ベースモデル	2005 年度から適用を開始, 2006 年度時点で適用件数は 20 件~30 件.
日本ユニシス	汎用, 開発全体の工数	FP か SLOC を規模パラメータとする COCOMO II 4) ベースのモデル	2000 年より COCOMO II 4) を含めた複数手法による見積もりを推奨. 利用実績は不明.
日本 IBM	汎用, 開発全体の工数	工数 = $\alpha \times \{10^{(\log \text{開発サイズ})} \beta\}^\gamma$ ※ $\alpha, \beta, \gamma$ は定数である.	開発した Web アプリケーションを社内ネットワークで提供. 12 年間運用し, 10,000 以上に適用.
ジャステック	汎用, 工程毎のアプリケーション開発工数	設計書文字数か SLOC を規模パラメータとし, 規模 $\times (1 + \Sigma 13$ の変動要因) $\times$ 生産性 $\times (1 + \Sigma 28$ の変動要因) で算出.	2003 年度, 同社における全プロジェクト (新規開発 150 件, 保守 95 件) に利用. 教育, 障害対応, 移行など (35 件) には未適用.
NTT データ ※本稿の内容	Web3 層構成システムが対象, 基盤構築工数	工数 = $\alpha \times \text{論理サーバ台数}^\beta \times \text{非機能要求グレード} \times \text{基盤構築期間}^\gamma$ ※ $\alpha, \beta, \gamma$ は定数, 詳細は 3 章で説明.	2008 年度から取組開始. Web アプリケーションを開発し, 社内ネットワークでグループ会社含めて提供. 2012 年度現在で利用者は 479 名.

※FP や SLOC などの規模を表すパラメータに, 規模あたりの開発工数 (文献 31)32)33)では「生産性」と呼ばれている)を乗じるだけの算出方法は, 見積もりモデルの利用事例ではないと判断した.

大幅な見積もり誤りを防止するため, NTT データでは 2008 年度よりシステム基盤構築の工数見積もりモデルについて継続的な改善とプロジェクトへの普及展開を進めている 14). システム基盤構築 (または単に基盤構築) は, ハードウェア, OS, ミドルウェア, ネットワークを組み合わせてアプリケーションソフトウェアを稼働させるための土台を作る作業であり, IT システム開発に要する工数の大きな部分を占める作業である. 図 1 に NTT データにおけるプロセス定義と, その中での基盤構築の位置付けを記す.

本稿では, 図中破線部の作業を基盤構築と呼び, それらの作業に要する工数の見積もりモデルについて述べる. NTT データでは, 開発工程にあたる概念をプロセス, 各プロセスで実施する作業をタスクと呼んで手順を定義している. 本稿の見積もりモデルでは, 図中破線部に含まれる基盤の詳細設計, 基盤制御機能開発, 環境構築, 運用ツール開発, 基盤のテスト, システムテスト (基盤部分) として定義されたタスクの工数を見積もる. 図 1 に示すように, これらのタスクは要件定義の後, アプリケーション開発, 既存システムからの移行準備, 運用手順書や体制整備, システムのオペレータの研修や体制整備などのサービス提供準備と並行して実施することが一般的である.

NTT データでは多くの場合, これらのタスクは基盤グループや方式担当などと呼ばれる基盤構築専任要員のチームが実施する. 特に大規模システム開発においては, 業務グループ, または AP 担当などと呼ばれるアプリケーション開発要員のチームと区別されて編成されることが多い. 本稿の取り組みで報告する見積もりモデルは, 基盤グループが作業毎の工数を積み上げ

て見積もった工数と比較することで, その妥当性を検証し, 大幅な見積もり誤りを防止することを目的としている.

本稿では, 工数見積もりモデルを企業のシステム開発実務で利用している事例として, NTT データにおける基盤工数見積もりモデルの継続的改善と普及展開について述べる. 取り組みの内容と実務者からのフィードバックについて報告することで, 見積もりモデルの実用化を検討している企業の実務者, PMO など支援組織の担当者, 企業と共同研究を行う学術機関の研究者に, 取り組みの進め方や期待される成果などの参考情報を提供する.

また本稿では, 作成した見積もりモデルの精度についても検証する. 基盤構築の工数見積もりの事例 19)20)37)41)はいくつか報告されているものの, 見積もりの精度が十分検証された事例は少ない. 本稿では, 取り組みの中で, モデル作成に使うデータやパラメータの変化に応じた見積もり精度の期待値の変化を, モデル作成に利用したデータの統計分析によって検証する. また, 見積もり精度の期待値を従来研究 24)と比較し, 見積もりモデルが実用に耐える精度を達成できたかどうかを検証する.

以降, 2 章で関連研究として, 企業における工数見積もりモデルの実用事例, 基盤工数見積もりの研究を紹介する. 3 章で NTT データにおける取り組みを紹介する. 4 章で実プロジェクトのデータを利用して見積もりモデルの精度の期待値を分析し, 取り組みに応じた年度毎の精度変化を検証する. 5 章で既存研究との精度比較によって実用性を考察し, 実務者からのフィードバックについて紹介する. 6 章で結論と今後の課題を述べる.

## 2. 関連研究

### 2.1 企業における工数見積もりモデルの実用事例

Molokken ら 24)は過去のサーベイ 2)11)21)23)40)に基づき、企業における工数見積もりモデルの利用を調査した。調査対象の国や対象者の役割、調査方法が異なるため結果を容易に一般化するのは難しいものの、いずれの調査でも最も多く利用されたのは専門家による判断 (expert judgment) であり、工数見積もりモデルを利用したという回答は多くなかった (全調査対象の 13%~26%であった)。Molokken らの調査によると、多くの企業で見積もりモデルが実用化されているとは言い難いと考えられる。

独立行政法人情報処理推進機構が日本のシステム開発企業における工数見積もりモデルの利用事例について紹介している 31)32)33)。文献 31)では新規開発における見積もりモデルの利用として表 1 の事例を紹介している。また、文献 32)では機能追加や維持保守開発、文献 33)ではテスト工程における見積もりモデルの利用を紹介している。日本 IBM やジャステックなど、ごく一部の企業では実務利用が進んでいる一方、多くの企業で広く長期にわたって利用が進んでいるとは言い難い。

表 1 の最下行に本稿で報告する NTT データにおける事例を併記する。文献 31)の事例と比較すると、基盤構築工数の見積もりモデルを実務利用している点が他と大きく異なる (文献 32)33)でも当該事例の紹介はない)。また、本稿の事例では、文献 31)32)33)では述べられていない見積もりモデルの有効性についても検証する。本稿の事例は稀少であり、参考情報としての価値は高いと考えられる。

### 2.2 システム基盤工数見積もりの研究

Valerdi37)は、システム開発全体の工数見積もりモデル COSYSMO (Constructive Systems Engineering Cost Model) を提案した。COSYSMO が見積もる工数には、基盤の設計やシステムテストの作業も含まれるが、基盤構築の大部分を占める基盤制御機能開発、環境構築といったタスクは含まれない。また、COSYSMO の見積もり対象は、システムの要件定義、マニュアル作成、移行作業など、基盤グループに割り当てられない可能性が高いタスクも含む。本稿では、基盤構築に要する作業の工数を必要十分に見積もるモデルの利用事例を報告する。

金子ら 19)20)は、大規模 IT インフラ構築工数の見積もりモデルを提案した。Logical Partitioning, High Availability Cluster Multiprocessing, IBM WebSphere Application Server, IBM Universal Database といった基盤構成要素数を規模と考えて工数を見積もる。本稿の見積もりモデルは、これらの基盤構成要素数と同様の概念である論理サーバ台数に加え、システムに求められる非機能要求の高さを表す非機能要求グレード 34)と開発期間も変数として工数を見積もる。それらのパラメータを考慮することで、見積もり精度の改善が期待できる。

小澤 41)は、サーバの種類と処理パターンから基盤構築工数を工程毎に見積もるモデルを提案した。典型的なサーバの種

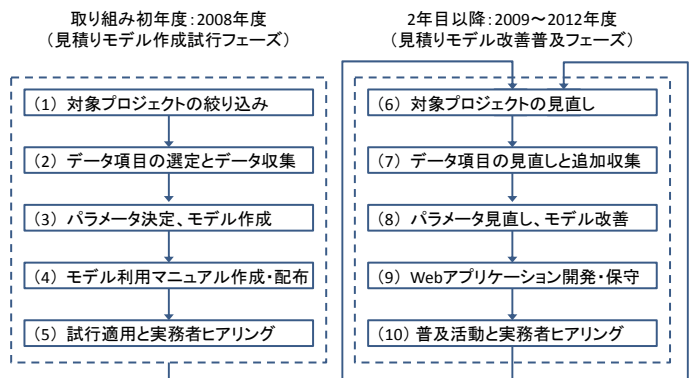


図 2 見積もりモデル作成と普及改善の手順

Figure 2 Procedure for Deriving and Improving Estimation Model

類と処理パターンについて実績データを分析し、難易度係数を決定する。構築する基盤のサーバ種類と処理パターンに対応する難易度係数から規模変数を算出し、3 種類の線形モデルに代入して、基盤設計・構築・テストの工数を各々見積もる。本稿の見積もりモデルも、小澤の方法と同様に、実績データを分析して係数を決定する。加えて本稿では、係数決定に利用した実績データを分析し、小澤の研究では報告されていない見積もりモデルの精度や実用性についても検証する。

## 3. NTT データにおける取り組み

図 2 に取り組みの手順を示す。2008 年度、基盤構築の大幅な見積もり誤りの発生を契機として研究部門が取り組みを開始した。2008 年度に見積もりモデルを作成し、文書化してデータ収集の協力者に配布した。2009 年度以降はモデルを改善し、Web アプリケーションに組み込んでサービスを提供した。図中の各手順では次を実施した。

### (1) 対象プロジェクトの絞り込み

実用に耐える精度を得るために見積もりモデルの対象とするプロジェクトを絞り込んだ。基盤構築の有識者と議論し、下記のプロジェクトを対象とした。

- **システム基盤を新規構築または完全更改**

既存の基盤を拡張する場合より工数が多く必要であり、不確実性が高い。大幅な見積もり誤りが生じる可能性が高く、見積もりモデルに対するニーズが大きいと思われる。システム基盤を新規構築または完全更改するプロジェクトを対象とした。

- **システム基盤のアーキテクチャが Web3 層構成**

2008 年度時点で主流であったオープンシステムによる Web3 層構成のアーキテクチャを対象とした。メインフレーム構成の基盤を新規構築することは稀であり、Hadoop などを利用する大規模クラスタ基盤、同様の基盤構成を仮想化して大量複製するクラウド基盤も、当時は稀であったため対象外とした。

- **システム基盤を既存ミドルウェアの組み合わせで構築**

性能要件が非常に厳しい場合など、ミドルウェアを個別

表 2 選定された収集データ項目

Table 2 Selected Data Items for Data Collection.

分類	項番	データ項目名	説明	影響の大きさ*
規模	1	論理サーバ台数	システムを構成するサーバを、物理・仮想を区別せずに 1 台と数える。クラスタを構成するサーバは個別に数える。	9.00
	2	クラスタ構成数	システム内に含まれているクラスタの数。	12.00
	3	クライアント数	システムが管理するクライアントの数。	7.00
	4	ネットワーク機器数	上記以外の設計が必要な機器(ファイアウォールなど)の数。	8.00
	5	外部接続先数	システムが接続する外部システムの数。	4.00
	6	分散拠点数	地理的に分散している拠点の数。	6.00
	7	パッケージプログラム数	基盤として利用しているパッケージプログラムの数。	3.00
	8	テーブル数	管理しているデータベース中のテーブルの数。	6.00
難易度	9	許容計画停止頻度	運用上、許容される計画停止の頻度。	9.00
	10	稼働率	運転時間に対する実際の稼働時間。	13.00
	11	サービス中断時間	システム切り替えによる中断時間。	6.00
	12	RPO(目標復旧時点), RTO(目標復旧時間)	障害発生時、データを復旧すべき時点。障害発生時、データの復旧に要する時間。	6.00
	13	ピーク時レスポンスタイム, 多重度, TPS	ピーク時のレスポンスタイム, 同時アクセス数, 1 秒あたりに処理すべきトランザクション処理数。	6.00
	14	縮退時レスポンスタイム, 多重度, TPS	縮退時のレスポンスタイム, 同時アクセス数, 1 秒あたりに処理すべきトランザクション処理数。	11.00
	15	帳票スループット性能	帳票の単位時間あたりの出力枚数。	10.00
	16	情報の機密性	システムが取り扱う情報の機密度。	6.00
	17	運用の自動化レベル	システム運用の自動化の度合い。	8.50
	18	許容移行停止時間	許容される移行のためのシステム停止時間。	12.00
	19	移行データ量・複雑さ	移行データの量および移行データの複雑さ。	6.00
開発能力	20	アーキテクチャの経験有無	類似アーキテクチャの構築経験有識者が体制にいるか否か。	1.00

※「影響の大きさ」に基盤構築工数への影響の大きさを示す。洗い出した全パラメータについて、7名の有識者に相対順位をアンケート調査した。1が最も影響が大きく、数字が小さいほど影響が大きいという回答を示す。上表には7名の回答の平均値を示す。

に開発・拡張する場合は対象外とした。基盤構築作業の一環として基盤制御のためのソフトウェアを開発・改良する場合、その部分の工数は本モデルとは別に計算する必要がある。

## (2) データ項目の選定とデータ収集

基盤構築に携わる有識者7名と議論し、基盤構築に影響を与えるパラメータを洗い出した。洗い出したパラメータを基盤の規模を表すもの、構築の難易度を表すもの、開発能力を表すものに分類し、影響の大きさを各有識者にアンケート調査した。影響が大きいと回答された表2の20パラメータを、収集するデータ項目として選定した。選定した項目について、2つの部署のプロジェクト10件からアンケート調査でデータを収集した。

## (3) パラメータ決定, モデル作成

収集したデータを分析してパラメータ決定し、モデルを作成した。収集したデータはプロジェクト10件であったため、式(1)に示すパラメータ2つの重回帰モデルを作成した。Cohenの効果量 $f^2$ などに基づく単回帰モデルを作成することが妥当である可能性もあったが、有識者との議論で基盤構築の難易度や構築能力も含めたいという要望が強く、統計的妥当性の低下を許容して規模以外に1つの補正パラメータを含めた。

$$\text{工数} = \alpha \times \text{規模パラメータ}^{\beta} \times \text{補正パラメータ} \quad (1)$$

パラメータ決定では、回帰モデルで算出した予測工数と各プロジェクトの実績工数から、式(2)で相対誤差(MRE:

表 3 見積もりモデルの係数

Table 3 Coefficients of Effort Estimation Model

係数		2008	2009-2011	2012
$\beta$		0.92 (1.00)	0.98 (1.14)	0.75 (1.43)
$\gamma$		—	—	0.80 (1.56)
アーキテク チャ経験	有	1.0	—	—
	無	4.37 (1.00)	—	—
非機能要求 グレード	低	—	1.0	1.0
	中	—	1.60 (1.29)	1.91 (1.22)
	高	—	5.25 (1.37)	3.99 (1.45)

※( )内に各係数に対する VIF の値を示す。また、—は当該年度のモデルにその係数が存在しないことを表す。

Magnitude of Relative Error) を算出し、全プロジェクトの中央値 (MdMRE: Median of MRE), MRE0.25 以下で見積もれたプロジェクトの割合 (Pred25) を基準とした。MRE の値域は[0.00, ∞)であり値が小さい程精度が高いことを、Pred25 の値域は[0.00, 1.00]で値が大きい程精度が高いことを表す。

$$MRE = \frac{|実績工数 - 予測工数|}{実績工数} \quad (2)$$

また、各パラメータの分散拡大要因 (VIF: Variance Inflation Factor) を算出し、作成したモデルに多重共線性が発生していないことを確認した。多重共線性はモデルのパラメータ同士が強く相関している状態であり、見積もり精度を大きく低下させる過適合の原因となる。VIF の値域は[0.00, ∞)であり、5 を超える場合は多重共線性が発生している可能性が高い[10]。

まず規模パラメータの候補 (表 2 の項番 1~8) 各々について式(1)の第 2 項、補正パラメータを除く単回帰モデルを作成し、精度が最も高くなる規模パラメータを選択した。次に、決定した規模パラメータの候補と補正パラメータの候補 (表 2 の項番 9~20) について式(1)の重回帰モデルを作成し、精度が最も高くなる補正パラメータを選択した。

2008 年度に作成したモデルを式(3)に示す。規模パラメータとして論理サーバ台数 (表 2 の項番 1), 補正パラメータとして構築対象アーキテクチャの経験有無 (表 2 の項番 20) を選択した。式中、 $\alpha$ ,  $\beta$ , 経験有無は係数を表す。機密情報のため  $\alpha$  は非公開とし、他の係数、および各係数の VIF を表 3 に示す。作成したモデルの MdMRE は 0.31, Pred25 は 40%であった。VIF はいずれも 1.00 であり、多重共線性発生の可能性は低いと考えられる。なお、アーキテクチャ経験有の場合に係数が 1.0 となるようダミー変数化したため、VIF は算出していない。

$$工数 = \alpha \times 論理サーバ台数^\beta \times 経験 \quad (3)$$

表 4 適用可能プロジェクトとプロジェクトデータ件数

Table 4 Applicable Projects and # of Projects in the Data

	2008	2009-2011	2012
構築種別	基盤を新規構築または完全更改		
アーキテクチャ	Web3 層構成		
構築方法	既存ミドルウェアの組み合わせ		
規模	3~15 台	3~30 台	3~70 台
構築期間	—	—	6~30 カ月
要員数	—	—	2~30 名
データ件数	10 件	33 件	43 件

※—は当該年度のモデルでは未定義であることを表す。

#### (4) モデル利用マニュアル作成・配布

作成した見積もりモデルについて、モデルの利用方法、意味、作成方法や分析結果をモデル利用マニュアルとして文書化して配布した。データ収集に協力してくれた 2 つの部署 10 のプロジェクトに電子メール送付すると共に、当該部署の幹部と管理職に説明会を実施した。

#### (5) 試行適用と実務者ヒアリング

データ収集に協力してくれた部署の実務者 5 名に作成したモデルを試行的に利用してもらい、ヒアリングでモデルの改善方法と普及展開方法を検討した。モデルのさらなる改善や、継続的な改善の必要性を示唆する意見も多かったが、それらも含め、取り組みについて肯定的な意見が大半であった。得たフィードバックの詳細は 5.2 に示す。

#### (6) 対象プロジェクトの見直し

データの分析結果および実務者から得たフィードバックに基づき、想定する見積もり対象を見直した。より多くのプロジェクトに利用してもらえるよう、表 4 のように、見直した結果を「適用可能プロジェクト」として見積もりモデルのマニュアルに明記した。

#### (7) データ項目の見直しと追加収集

表 2 に示すデータ項目を、見積もり精度への寄与、利用時の入力の手軽さなどに基づいて見直した。有識者との議論に基づき、Putnam モデル [28] に倣い、基盤構築期間を追加収集することとした。また、2008 年度当初は表 2 の項番 9~19 の項目で基盤構築の難易度を収集していたが、項目同士の相関が高く、多重共線性の原因となるため、非機能要求グレード [34] という 1 つの項目に統合して次の 3 段階で収集した。

- ・低：システムトラブルの影響が特定の組織や企業内に閉じており、社会や業界まで広がる可能性が低いシステム。公共、法人分野の非基幹系システムを想定。
- ・中：システムトラブルの影響が社会、または特定業界の一部に広がり、悪影響を及ぼすシステム。地方自治体のシステムや公共法人分野の基幹システムを想定。
- ・高：システムトラブルが社会、または特定業界全体に甚

表 6 実施した普及展開とユーザ数の増加

Table 6 Implemented Promotions and Growth Number of

	Users				
	2008	2009	2010	2011	2012
個別説明	8	12	3	0	3
利用説明会	2	2	3	0	0
社内セミナー	1	1	2	5	9
社内展示会	0	1	0	0	1
バージョンアップ	1	1	1	0	6
バージョンアップ 周知	0	1	0	0	2
読み物記事	0	0	0	0	3
年間ユーザ数*	5	48	102	193	479

※2008 年度は試行利用した実務者 5 名を、2009 年度以降は Web アプリケーションの月次ユニークユーザの年度計を記す。

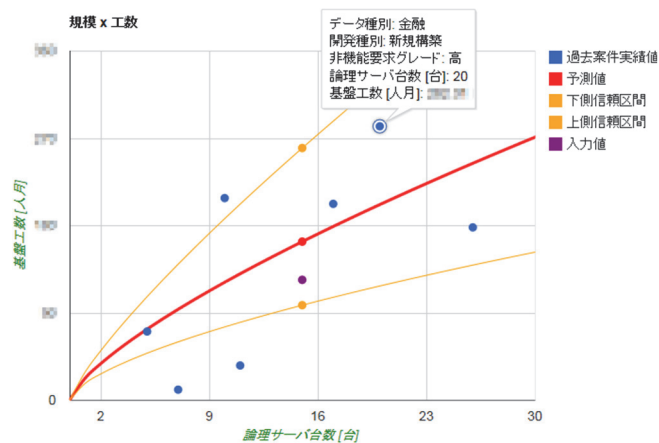


図 3 Web アプリケーションの見積もり結果表示画面例  
 Figure 3 An Estimation Result Displayed on the Web Application  
 ※図中、縦軸の工数は機密情報のため、モザイク化した。

大な悪影響を及ぼすシステム。金融分野のミッションクリティカルシステムや公共分野の中央省庁システムを想定。

これらのデータ項目を、年間 5 件～10 件程度のプロジェクトについて継続的にヒアリングで収集した。表 4 に示すように、2008 年当初 2 つの部署のプロジェクト 10 件であったデータは、2012 年度には 20 の部署のプロジェクト 43 件に増加した。

(8) パラメータ見直し、モデル改善

データの収集状況、有識者や利用者との議論に基づき、パラメータを見直し、モデルを修正した。2009～2011 年度に収集したデータを追加して修正したモデルを式(4)、さらに 2012 年度に収集したデータを追加して修正したモデルを式(5)に示す。

$$\text{工数} = \alpha \times \text{論理サーバ台数}^{\beta} \times \text{非機能要求グレード} \quad (4)$$

表 5 検証に利用したデータ

Table 5 Data Used for The Evaluation

		2008	2009-2011	2012
データ件数		10 件	33 件	43 件
基盤構築工数 (人月) [比率尺度]	変動係数	1.15	2.54	2.66
	論理サーバ台数 (台) [比率尺度]	26.00	100.00	100.00
論理サーバ台数 (台) [比率尺度]	中央値	8.00	10.00	10.00
	最小値	3.00	2.00	2.00
	変動係数	0.73	1.18	1.15
	基盤構築期間 (月) [比率尺度]	23.00	28.00	28.00
基盤構築期間 (月) [比率尺度]	中央値	11.00	12.00	11.00
	最小値	4.00	4.00	4.00
	変動係数	0.44	0.50	0.50
アーキテクチャ経験有無 [名義尺度]	有/無			
非機能要求グレード [順序尺度]	低/中/高			

※ [ ] 内に尺度水準を記す。機密情報のため、基盤構築工数については変動係数のみを示す。変動係数は標準偏差を平均値で除算した値であり、相対的なばらつきを度を示す。

$$\text{工数} = \alpha \times \text{論理サーバ台数}^{\beta} \times \text{非機能要求グレード} \times \text{開発期間}^{\gamma} \quad (5)$$

式(4), (5)では、主観的に判断される「アーキテクチャ経験」より、客観的に判断できるパラメータの方が、利用しやすいという利用者の意見を受け、補正パラメータを「非機能要求グレード」に変更した。また、式(5)ではさらに「開発期間」を追加した。機密情報のため  $\alpha$  は非公開とし、他の係数、および各係数の VIF を表 3 に示す。2009～2011 年度モデルの MdMRE は 0.46, Pred25 は 33%であった。また、2012 年度モデルの MdMRE は 0.34, Pred25 は 44%であった。VIF は最も大きい場合でも 1.56 であり、多重共線性発生の可能性は低いと考えられる。なお、非機能要求グレード低の場合に係数が 1.0 となるようダミー変数化したため、VIF は算出してない。

(9) Web アプリケーション開発・保守

2009 年度に見積もりモデルを組み込んだ Web アプリケーションを開発した。アプリケーションの見積もり結果表示画面例を図 3 に示す。横軸に論理サーバ台数 (台)、縦軸に基盤構築工数 (人月) をとる散布図上に、入力パラメータに対する見積もり結果と見積もり結果の 95%信頼区間を折れ線グラフで表示する。また、実施予定プロジェクトの計画工数と、モデル作成に利用した過去プロジェクトを

表 7 精度変化の検証結果 (MdmRE)

Table 7 Evaluation Result of Accuracy Changes with MdmRE

データ件数		2008	2009-2011	2012
		10 件	33 件	43 件
2008	式(3)	0.31	<u>(0.08)</u> 0.57	(0.65) 0.55
2009-2011	式(4)	(0.70) 0.28	(0.34) (0.15) 0.46	<u>(0.06)</u> (0.14) 0.34
2012	式(5)	(0.43) 0.27	(0.97) <u>(0.02)</u>	(0.28) <u>(0.10)</u> (0.80) 0.34

※各セルの値の左, 上, 左上に記述した()内に, 左, 上, 左上の各セルの値との差の検定結果 (p 値) を示す.  $p \leq 0.10$  の場合, 危険率 10% で有意差があると判断し, p 値を下線で強調する.

異なる色でプロット表示し, プロットにマウスポインタを重ねると図 3 のような情報を表示する. ユーザはプロジェクトの計画工数を, 見積もり結果を表す折れ線グラフや過去プロジェクトと比較することで, その妥当性を検証できる.

#### (10) 普及展開と実務者ヒアリング

実施した普及展開活動の頻度とユーザ数の増加を表 6 に示す. 2008 年度~2009 年度当初は, 主に数人の実務者や開発チームに対する個別説明, または部署の管理職や課長代理層十数名に対する説明会による普及展開を実施した. 2010 年度からは, より大人数にアピールできる社内セミナーや社内展示会を企画・実施した. 2012 年度には Web サイトに読み物記事を掲載し, その読み物記事や公開している Web アプリケーションへの誘導記事を社内のポータルサイトに掲載することで, ユーザ数の大幅増加に成功した. 2008 年度 5 名だったモデルの年間利用者数は, 2012 年度には合計 479 名まで増加した.

普及展開の過程で実務者に見積もりモデルや Web アプリケーションについてヒアリングを行い, 適宜フィードバックを収集した. 収集したフィードバックに基づいて, 取り組みや普及展開の方針を決定・変更すると共に, 継続的に Web アプリケーションを改善した. Web アプリケーションに大きな機能追加を行った場合は, 社内広報システムを通じて国内全グループ会社に周知し, 新規ユーザへの宣伝活動を兼ねてバージョンアップを通知した. 実務者から収集したフィードバックの詳細を 5.2 に, バージョンアップとバージョンアップ周知の頻度を表 6 に示す. 社内セミナーや読み物記事の掲載に加え, これらの社内周知もユーザ数の向上に寄与したと考えられる.

## 4. 見積もりモデルの精度変化の検証

### 4.1 概要

取り組みによる見積もりモデルの精度変化について検証するため, モデル作成に利用した表 5 のデータを分析し, データおよびモデルパラメータを 2008 年度, 2009~2011 年度, 2012 年度の状況に各々変更して MdmRE と Pred25

表 8 精度変化の検証結果 (Pred25)

Table 8 Evaluation Result of Accuracy Changes with Pred25

データ件数		2008	2009-2011	2012
		10 件	33 件	43 件
2008	式(3)	40%	(0.21) 18%	(0.78) 23%
2009-2011	式(4)	(1.00) 40%	(0.71) (0.26) 33%	(0.13) (0.34) 35%
2012	式(5)	(1.00) 40%	(0.72) (0.21) 52%	(0.36) (0.51) 44%

を計算した. データについては, 表 5 に示すとおり各年度でモデル作成に利用した 10 件, 33 件, 43 件を学習データおよびテストデータとした. モデルパラメータは各年度に応じて式(3)~(5)のように変更した. データ追加収集による精度の変化と, パラメータ見直しによる精度変化を評価するため, 表 7, 表 8 に示すようにパラメータを固定してデータを変化させた場合, データを固定してパラメータを変化させた場合, データとパラメータいずれもを変化させた場合とで計 9 回 MdmRE と Pred25 を算出して比較した.

### 4.2 データ追加収集による精度変化

データ追加収集による精度変化を表 7, 表 8 の列方向に示す. モデルパラメータを固定し, データを変化させた場合の精度変化を, 左から右へ古い年度から順に記す. 各年度の値について, MdmRE では Wilcoxon の順位和検定 25), Pred25 では Fisher の正確確率検定 25) で前年度の値との差を検定した. 各セルの値の左側に記す()内に, 直左のセルに記す前年度の値との差の検定結果 (p 値) を示す.  $p \leq 0.10$  の場合, 危険率 10% で有意差があると判断し, p 値を下線で強調する.

例えば, 表 7 の最上行中央列のセルは, 2009~2011 年度のデータ 33 件を分析し, 2008 年度のモデルパラメータで式(3)の重回帰モデルを作成した結果である. 0.57 は, 33 件のプロジェクトについて算出した MdmRE である. 当該セルの値の左に記す(0.08)は, 2008 年度のデータ 10 件を分析した場合の MdmRE 0.31 との差の検定結果 (p 値) である. ここでは,  $p \leq 0.10$  であり, 危険率 10% で有意差があると判断できるため, 0.08 を下線で強調している.

検証の結果, モデルパラメータを 2008 年度に固定した場合, 2009~2011 年度の大幅なデータ増加で MdmRE, Pred25 共に精度が大きく悪化した (MdmRE は統計的有意差も観察できた). この理由として, 2009~2011 年度の追加収集データに, 2008 年度より多様なプロジェクトが含まれたことが考えられる. 表 5 に示すように, 2009~2011 年度のデータは 2008 年度に比べて基盤構築工数と論理サー

バ台数の最大値と変動係数が大きく増加している。これにより、3章(6)で説明したように適用可能プロジェクトを拡大できた一方、モデルパラメータを見直さなければ、精度が大きく悪化してしまったと考えられる。

2012年度のデータ追加収集は、2009～2011年度ほど大きな精度の変化はなく、MdmRE, Pred25共に統計的有意差は観察できなかった。表5に示すように、基盤構築工数、論理サーバ台数、基盤構築期間いずれも最大値や変動係数に前年度からの大きな変化はない。2012年度は、データが10件程度増加したが、データに含まれるプロジェクトの多様性は大きく変化しなかったと考えられる。

#### 4.3 パラメータ見直しによる精度変化

パラメータ見直しによる精度変化を表7、表8の行方向に示す。学習データおよびテストデータを固定し、モデルパラメータを変化させた場合の精度変化を、上から下へ古い年度から順に記述する。各年度の値について、MdmREではWilcoxonの符号順位検定(25)、Pred25ではFisherの正確確率検定(25)で前年度の値との差を検定した。各セルの値の上に記す( )内に、直上のセルに記す前年度の値との差の検定結果( $p$ 値)を示す。 $p \leq 0.10$ の場合、危険率10%で有意差があると判断し、 $p$ 値を下線で強調する。

検証の結果、データを固定した場合、全ての年度のデータにおいて、多くの場合にMdmRE, Pred25共に精度が改善した。特に、2009～2011年度から2012年度のパラメータに変更した場合にはMdmREで統計的有意差も観察できた。2008年度から2009～2011年度のパラメータ変更(式(3)から式(4))では、2値のカテゴリ変数であるアーキテクチャ経験有無から、3値の非機能要求グレードに変更している。さらに、2012年度のパラメータ変更(式(4)から式(5))では比率尺度の開発期間を追加している。モデルが含む情報量の増加に従って精度が改善し、また3章(8)で説明した多重共線性の発生もない。パラメータを適宜適切に見直すことで、継続的に精度を改善できたと考えられる。

#### 4.4 データ追加収集とパラメータ見直しによる精度変化

データ追加収集とパラメータ見直しによる精度変化を表7、表8の斜め方向に示す。学習データおよびテストデータとモデルパラメータの両方を変化させた場合の精度変化を、左上から右下へ古い年度から順に記述する。各年度の値について、MdmREではWilcoxonの順位検定(25)、Pred25ではFisherの正確確率検定(25)で前年度の値との差を検定した。各セルの値の左上に記す( )内に、直左上のセルに記す前年度の値との差の検定結果( $p$ 値)を示す。 $p \leq 0.10$ の場合、危険率10%で有意差があると判断し、 $p$ 値を下線で強調する。

検証の結果、統計的に有意差は観察できなかったが、2009～2011年度はMdmRE, Pred25共に精度が悪化し、2012

表9 既存研究との見積もり精度の比較

Table 9 Comparison of Estimation Accuracy with Past

Researches				
	見積対象	データ件数	MRE	Pred25
2008	基盤構築 工数	10	(中央値) 0.31	40%
2009-2011		33	(中央値) 0.46	33%
2012		43	(中央値) 0.34	44%
金子ら 19)		5	(中央値) 0.02	60%
Valerdi37)	システム 開発工数	27	—	27%
Jenkinsら 16)	アプリケ ーション 開発工数	72	(中央値) 0.34	—
Phan27)		191	(平均値) 0.33	—
Bergeronら 2)		89	(平均値) 0.33	—
Standish36)		365	(平均値) 0.89	約20%

※—は、当該指標や実験結果データの報告がないことを示す。

年度は改善した。この理由として、表5に示すように、2009～2011年度のデータは2008年度に比べて最大値や変動係数が大きく増加したが、モデルパラメータは式(3)から式(4)のように実質的にはダミー変数を1つ追加しただけであったことが挙げられる。データに含まれるプロジェクトの多様性増大に対して、重回帰モデルの見積もり精度改善が小さかったため、2009～2011年度は精度が低下したと考えられる。一方、2012年度のデータは表5に示すように最大値や変動係数の増加はほぼなかったのに対し、式(5)のように比率尺度のパラメータを追加したため、精度が改善したと考えられる。

## 5. 考察

### 5.1 既存研究との見積もり精度比較による実用性の考察

見積もりモデルおよびツールの実用性を考察するため、既存研究で報告された精度との比較結果を表9に示す。表中、最上行から下へ順に、当取り組みにおける各年度のモデル、2.2節で紹介した金子ら19)とValerdi37)のモデル、2.1節で紹介したMolokkenら24)の調査で対象とされた文献16), 27), 2), 36)で報告された精度をデータ件数と共に示す。なお、2.2節で紹介した小澤41)の研究やMolokkenらの調査における他の研究では、精度が報告されていない



め比較できなかった。

比較の結果、2012年度に見積もりモデルを実用に耐え得る精度に改善できたと考えられる。表9に示す文献の多くでは、MREの中央値または平均値は0.33~0.34であり、Pred25は20%~27%であった。当取り組みのモデルは2012年度にある程度のデータ件数によって統計的妥当性を担保した上で、これと同程度の精度を達成できた。金子らの研究と比べると精度が低く見えるが、金子らの研究ではデータ件数が非常に少なく、過適合の発生など統計モデルとしての妥当性には疑問がある。一方、当取り組みでは、2012年度には、統計モデルとしての妥当性を十分担保できるデータ件数を確保できている。

## 5.2 実務者からのフィードバック

当取り組みによって作成した見積もりモデルやツールを利用した実務者から得られた典型的なフィードバックを次に示す。

- FP法1)など、アプリケーション開発の見積り方法は多く存在するが、基盤側の見積り方法は存在しないため有用である。
- 基盤構築工数の妥当性を顧客に説明する際に利用できた。
- 作業積み上げて見積もった工数と、モデルの見積もりが大きく変わらないことに驚いた。見積りの妥当性検証に利用できた。
- 設計の共通化や再利用による工数削減効果（生産性向上効果）を、モデルの見積もり結果と比較することで検証できた。
- 論理サーバ台数だけでなく、他の構成要素（ネットワークスイッチの数など）も考慮すれば精度が向上すると考える。
- 基盤の仮想化、再利用、自動化ツール利用により、基盤構築の生産性が近年大きく向上している。データ収集と再分析を継続実施し、モデル改善を続けて欲しい。

モデルや取り組みの継続的な改善が必要だとする意見も多かったが、それらも含めて取り組みに対して肯定的な意見が大半であった。特に、基盤構築に関わる実務者からは、アプリケーション開発の見積りについては多くの研究が実施されている(1)4)6)9)15)28)30)39)など)のに対し、基盤構築の手法や研究は少ないため、肯定的な意見が多かった。分析に必要なデータの収集についても、協力的であり、継続的な取り組みの実施と改善を求める意見が大半であった。2009年度以降、収集範囲を拡大できたのは、前年度の見積もりモデルをWebアプリケーションという利用しやすい形態で提供しつつ、適用可能プロジェクトの拡大というメリットを訴求できたためだと考えられる。

## 6. おわりに

本稿では、見積もりモデルを企業のシステム開発実務で利用している事例として、NTTデータにおける基盤構築工数見積もりモデルの継続的改善と普及展開について報告した。2008~2012年度の取り組みを紹介し、取り組みによる精度の改善について実プロジェクトのデータを利用して検証した。また、モデルの精度を既存研究と比較して実用性を考察し、実務者から得たフィードバックを紹介した。

2014年度現在、本稿で報告した基盤構築工数見積もりモデルに加えてアプリケーション開発工数見積もりモデルを作成し、両者を1つのWebアプリケーションに組み込んで国内の全グループ会社にサービス提供している。全グループ会社の社員は誰でも当該サービスを利用可能であり、大幅な見積もり誤りがないかの検証や、生産性向上効果の検証に利用されている。また、見積り誤りに起因する問題化プロジェクト発生防止のための第三者見積り審査への応用も検討されている。

今後の課題として、SAPなどのパッケージ製品を利用した開発、Hadoopなどを利用する大規模クラスタ基盤、仮想化技術を利用するクラウド基盤など、技術的に多様化する基盤構築方法に対応するため、それらの構築プロジェクトからデータを収集し、モデルを改善すること。また、国内だけでなく海外グループ会社からもデータを収集し、適用可能プロジェクトの拡大に応じてモデルパラメータを継続的に見直すと共に、開発したWebアプリケーションによるサービスを長期的な戦略の下に改善して行くことが挙げられる。

## 参考文献

- 1) Albrecht, A. J., and Gaffney, J. E.: Software Function, Source Lines of Code, and Development Effort Prediction, IEEE Trans. on Soft. Eng., Vol.9, pp.639-648 (1983).
- 2) Bergeron, F., and St-Arnaud, J.-Y.: Estimation of Information Systems Development Efforts: A Pilot Study, Information & Management, Vol.22, pp.239-254 (1992).
- 3) Blattberg, R. C., and Hoch, S. J.: Database Models and Managerial Intuition: 50% Model + 50% Manager, Management Science, Vol.36, No.8, pp.887-899 (1990).
- 4) Boehm, Barry W., Ray Madachy, and Bert Steece: Software Cost Estimation with Cocomo II, Prentice Hall PTR (2000).
- 5) Braun, P. A. and Yaniv, I.: A Case Study of Expert Judgment: Economists' Probabilities vs Base-Rate Model Forecasts, Journal of Behavioral Decision Making, Vol.5, No.3, pp.217-231 (1992).
- 6) Briand, L. C., Basili, V. R., and Thomas, W. M.: A Pattern Recognition Approach for Software Engineering Data Analysis, IEEE Trans. on Software Engineering, Vol.18, pp.931-942 (1992).
- 7) Cohen, J.: Statistical Power analysis for the Behavioral Sciences, 2nd ed., Hillsdale, NJ: Lawrence Erlbaum, (1988).
- 8) Connolly, T., and Dean, D.: Decomposed versus Holistic Estimates of Effort Required for Software Writing Tasks, Management Science, Vol.43, No.7, pp.1029-1045 (1997).
- 9) Delaney, W. A.: Predicting the Costs of Computer Programs, Data Processing Magazine, pp.32-34 (1966).

- 10) Dobson, A. J. (著), 田中 豊, 森川 敏彦, 山中 竹春, 富田 誠 (訳): 一般化線形モデル入門, pp.115-116, 共立出版 (2008).
- 11) Heemstra, F.J., and Kusters, R.J.: Controlling Software Development Costs: A Field Study, In International Conference on Organisation and Information Systems, Bled, Yugoslavia (1989).
- 12) Hill, J., Thomas, L. C., and Allen, D. E.: Experts' Estimates of Task Durations in Software Development Projects, International Journal of Project Management, Vol.18, No.1, pp.13-21 (2000).
- 13) Hoch, S. J., and Schkade, D. A.: A Psychological Approach to Decision Support Systems, Management Science, Vol.42, No.1, pp.51-64 (1996).
- 14) 井ノ口 伸人, 伏田 享平, 渡辺 絢子, 大杉 直樹, 藤貫 美佐, 渡辺 真太郎, 戸村 元久, 木谷 強: システム基盤構築の工数見積もりモデル・ツール開発と普及展開, ソフトウェア品質シンポジウム 2012 (2012).
- 15) Jørgensen, M.: Experience with the Accuracy of Software Maintenance Task Effort Prediction Models, IEEE Trans. on Soft. Eng., Vol.21, Issue.8, pp.674-681 (1995).
- 16) Jenkins, A.M., Naumann, J.D., and Wetherbe, J.C.: Empirical Investigation of Systems Development Practices and Results, Information & Management, Vol.7, pp.73-82 (1984).
- 17) Jørgensen, M.: A Review of Studies on Expert Estimation of Software Development Effort, Journal of Systems and Software, Vol.70, Issues.1-2, pp.37-60 (2004).
- 18) Kahneman, D., Slovic, P., and Tversky, A.: Judgment Under Uncertainty: Heuristics and Biases, Cambridge, United Kingdom, Cambridge University Press (1982).
- 19) 金子 美和, 伊藤 穰, 田中 正美, 弘末 太郎: 大規模 IT インフラ構築プロジェクトにおける概算工数見積もりモデル, PM 学会研究発表大会予稿集 2008(春季), pp.111-114 (2008).
- 20) Kaneko, M.: Model in Approximate Estimating Effort for Large IT Infrastructure Building Projects, In *Proc. of the ProMAC 2010*, pp.1301-1308 (2010).
- 21) Lederer, A. L., and Prasad, J.: Nine Management Guidelines for Better Cost Estimating, Communications of the ACM, Vol.35, No.2, pp.51-59 (1992).
- 22) Lim, J. S. and O'Connor, M.: Judgmental Forecasting with Time Series and Causal Information, International Journal of Forecasting, Vol.12, No.1, pp.139-153 (1996).
- 23) McAulay, K. Information Systems Development and the Changing Role of MIS in the Organisation, In First New Zealand MIS Management Conference, Wellington (1987).
- 24) Moløkken, K., and Jørgensen, M.: A Review of Surveys on Software Effort Estimation, In *Proc of the 2003 Intl. Symposium on Empirical Soft. Eng. (ISESE 2003)*, pp.223-230 (2003).
- 25) Myles, P. S., Gin, T.: Statistical Methods for Anaesthesia and Intensive Care, Butterworth-Heinemann, (2000).
- 26) O'Connor, M., Remus, W., and Griggs, K.: Judgmental Forecasting in Times of Change, International Journal of Forecasting, Vol.9, No 2, pp.163-172 (1993).
- 27) Phan, D.: Information Systems Project Management: an Integrated Resource Planning Perspective Model, In Department of Management and Information Systems, Arizona: Tucson (1990).
- 28) Putnam, L. H.: A General Empirical Solution to the Macro Software Sizing and Estimation Problem, IEEE Trans. on Soft. Eng., Vol.4, pp.345-381 (1978).
- 29) Shanteau, J.: Competence in Experts: The Role of Task Characteristics, Organizational Behaviour and Human Decision Processes, Vol.53, No.2, pp.252-266 (1992).
- 30) Shepperd, M., Shfield, C.: Estimating Software Project Effort Using Analogies, IEEE Trans. on Soft. Eng., Vol.21, pp.126-137 (1997).
- 31) 独立行政法人情報処理推進機構ソフトウェア・エンジニアリング・センター: ソフトウェア開発見積もりガイドブック～IT ユーザとベンダにおける定量的見積もりの実現～ (2006).
- 32) 独立行政法人情報処理推進機構ソフトウェア・エンジニアリング・センター: ソフトウェア改良開発見積もりガイドブック～既存システムがある場合の開発～ (2007).
- 33) 独立行政法人情報処理推進機構ソフトウェア・エンジニアリング・センター: ソフトウェアテスト見積もりガイドブック～品質要件に応じた見積もりとは～ (2008).
- 34) 独立行政法人情報処理推進機構ソフトウェア・エンジニアリング・センター: 非機能要求の見える化と確認の手段を実現する「非機能要求グレード」の公開 (2010).
- 35) Soll, J. B.: Determinants of Overconfidence and Miscalibration: The Roles of Random Error and Ecological Structure, Organizational Behaviour and Human Decision Processes, Vol.65, No.2, pp.117-137 (1996).
- 36) Standish Group: The Chaos Report, The Standish Group (1994).
- 37) Valerdi, R.: The Constructive Systems Engineering Cost Model (COSYSMO): Quantifying the Costs of Systems Engineering Effort in Complex Systems, VDM Verlag (2008).
- 38) Whitecotton, S. M., Sanders, D. E., and Norris, K. B.: Improving Predictive Accuracy with A Combination of Human Intuition and Mechanical Decision Aids, Organizational Behaviour and Human Decision Processes, Vol.76, No3, pp.325-348 (1998).
- 39) Wolverton, R. W.: The Cost of Developing Large-Scale Software, IEEE Trans. on Soft. Eng., Vol.C-23, Issue.6, pp.615-636 (1974).
- 40) Wydenbach, G, and Paynter, J.: Software Project Estimation: a Survey of Practices in New Zealand, New Zealand Journal of Computing, Vol.6, pp.317-327 (1995).
- 41) 小澤 良男: システム基盤設計構築における見積もり手法の確立, NRI 技術創発, Vol.9, pp.70-85 (2007).