付録 1．変化度の山を明瞭に示す w 値の観察データ

プロジェクト A，B に対して，CDAM による分析を行うため，SSA のパラメータ w について，4～80 に変化させながら各グラフを観察した．以下に観察した結果，特徴的であるものを示す．なお，グラフの棒グラフが検知した欠陥数の実数，折れ線グラフが変化度を示す．

① w=4，5 の場合は，変化度が10^{-16}~10^{-15}などほとんど 0 に近い値をとる．振幅は小さいが常に激しく振動している状態にあり，変化度の山を明瞭には判断することができない．そのため，不適であると判断する．

② w=5 を超えると変化度の振幅が 0.1 を超え，1～0 のスコープで観察可能な範囲になるが，変化の激しい傾向は変わらず，かなりの範囲（全体の半分以上）を変化度の山として捉えてしまうため，不適と判断する．

③ w=12（およそ四半期）を境に変化度の山を示す範囲が顕著に少なくなる．プロジェクト A の場合，図 3，および図 4 を見ると変化度の山を示す箇所が急減していることが分かる．プロジェクト B においては，図 17，図 18 が示すように w=10 を境に変化している．両方のプロジェクトを考慮すると w=12 の近傍から傾向が変化し，変化度の山を示す箇所が明瞭に示されるようになる．

④ w=12 を超えて増やしていくと，四半期の倍数（半期，年間など）においては変化度の山とそうでない値が明瞭になる．

⑤ w=26 を超えると，データ取得後 w に比例したタイミングで一度高い変化度を示す．それ以降についてはだんだんと変化度のメリハリが失われる傾向になる．

⑥ 上記から，変化度の山を明確に，かつ最初の山をできるだけ早く捉えることのできる，w=12 を分析に採用した．

上記①～⑤を読み取れるように w=4（1 か月，最小値），w=8（2 か月），w=11，12，13（3 か月，およびその近傍値），w=26（6 か月），w=52（1 年），w=78（1 年半）を指定し，特異スペクトル解析を行ったグラフを図 1～図 16 に図示する．また，プロジェクト B の w=9，w=10 のグラフを図 17，図 18 にそれぞれ示す．

プロジェクトAに関するグラフ

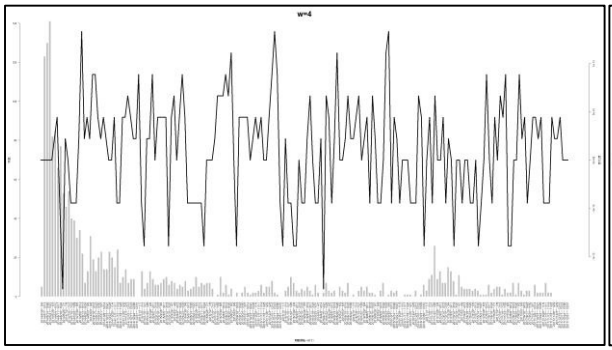

図1 プロジェクトA（w=4）

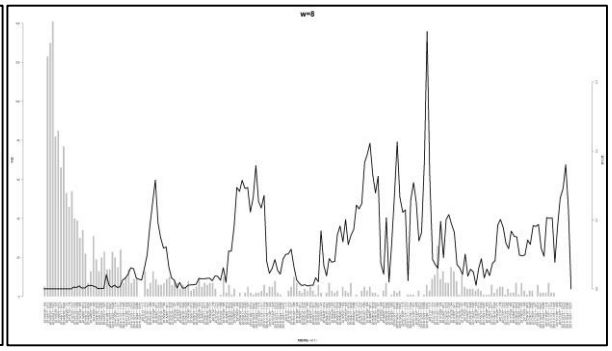

図2 プロジェクトA（w=8）

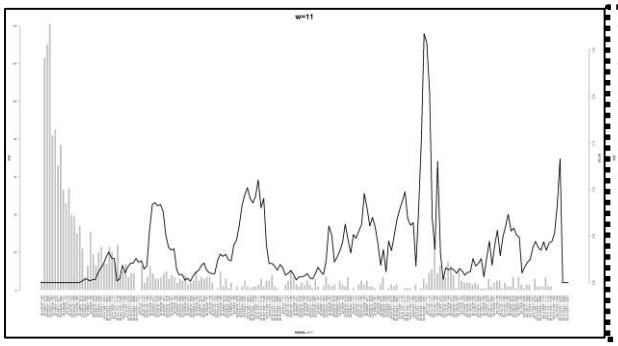

図3 プロジェクトA（w=11）

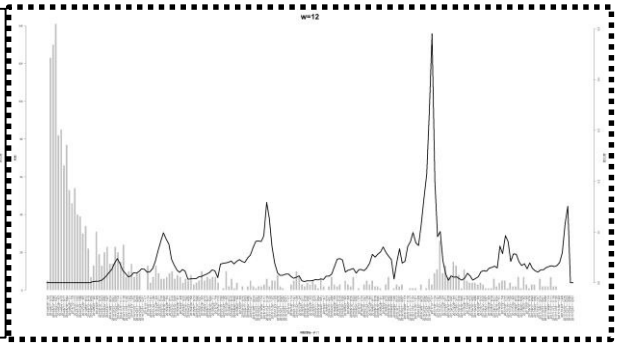

図4 プロジェクトA（w=12）

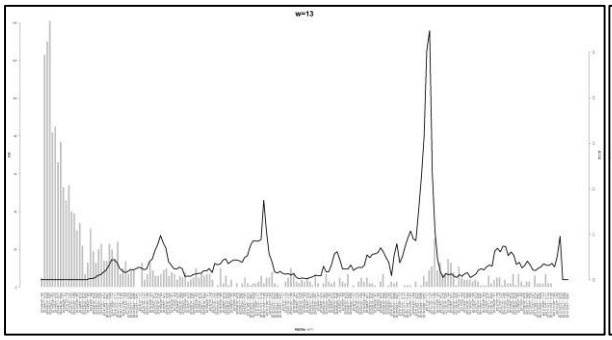

図5 プロジェクトA（w=13）

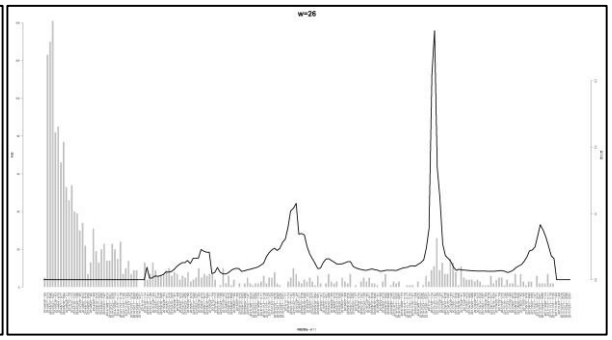

図6 プロジェクトA（w=26）

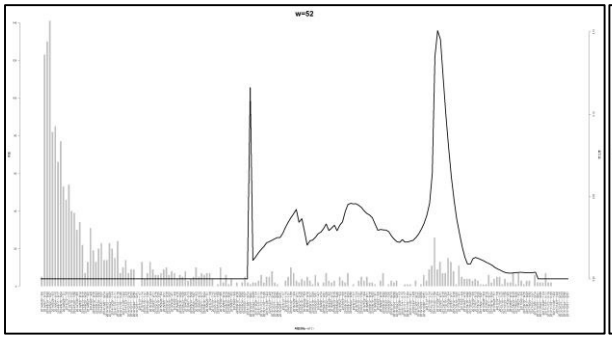

図7 プロジェクトA（w=52）

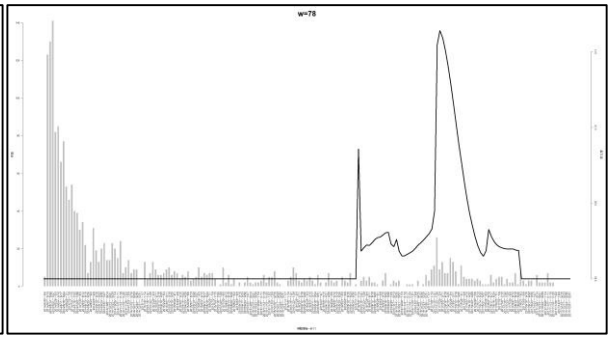

図8 プロジェクトA（w=78）

プロジェクトBに関するグラフ

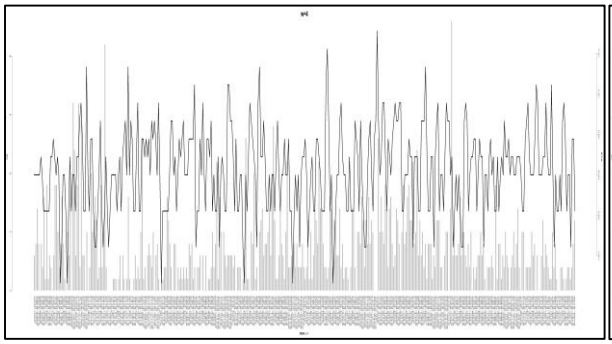

図 9 プロジェクトB（w=4）

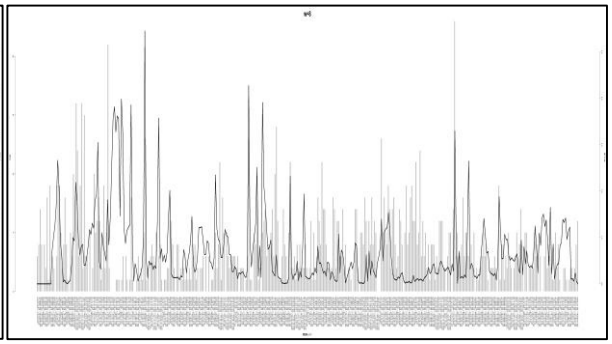

図 10 プロジェクトB（w=8）

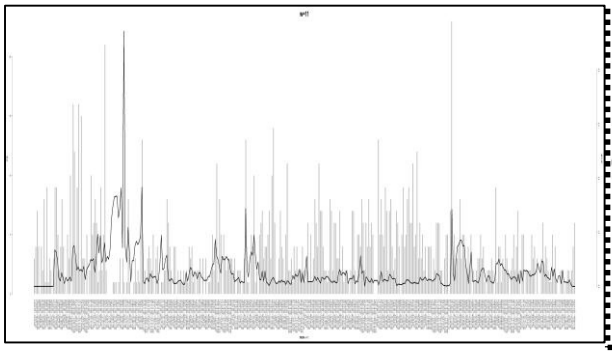

図 11 プロジェクトB（w=11）

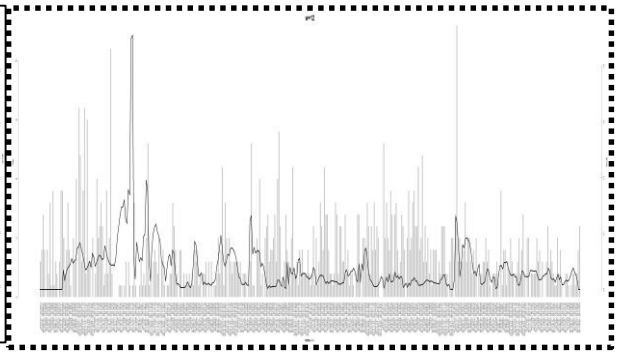

図 12 プロジェクトB（w=12）

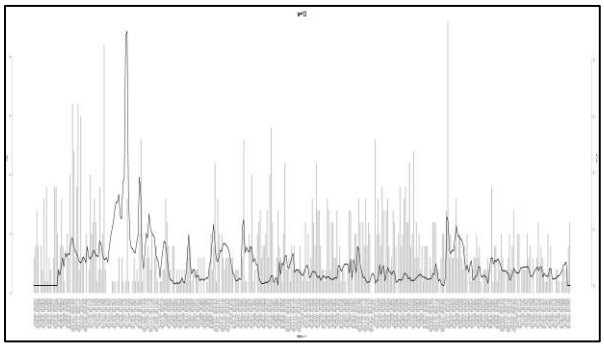

図 13 プロジェクトB（w=13）

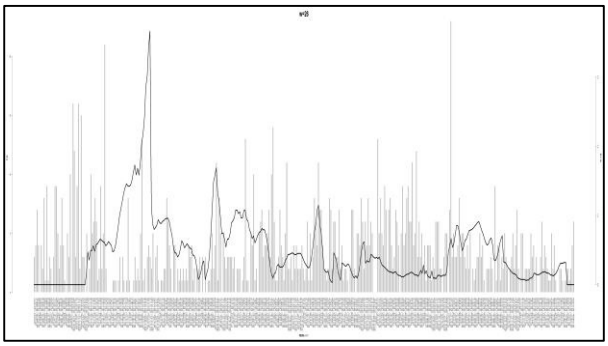

図 14 プロジェクトB（w=26）

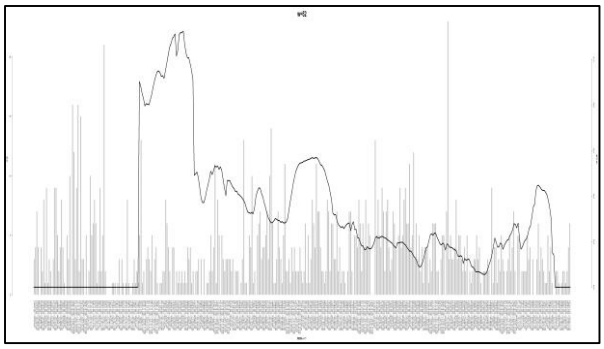

図 15 プロジェクトB（w=52）

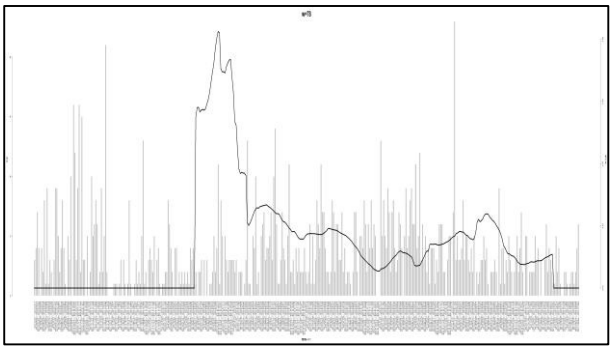

図 16 プロジェクトB（w=78）

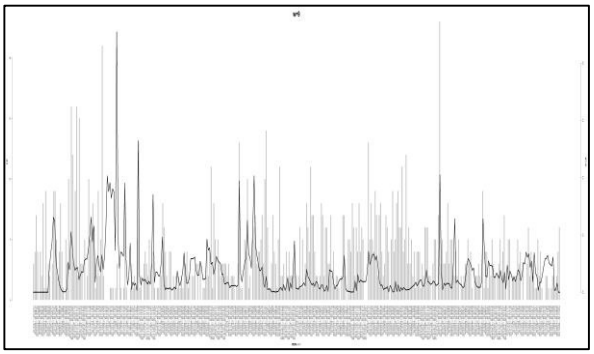

図 17 プロジェクト B（w=9）

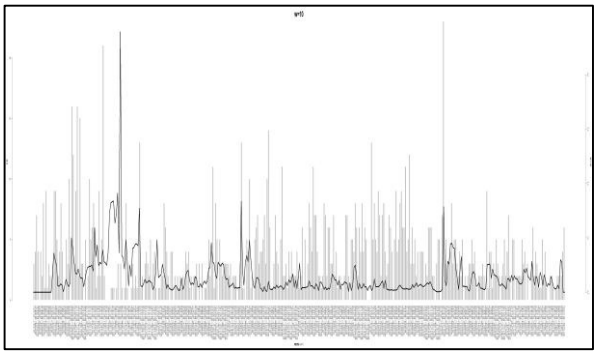

図 18 プロジェクト B（w=10）

付録 2. 変化度に基づく欠陥検知傾向の変化とプロジェクト状況

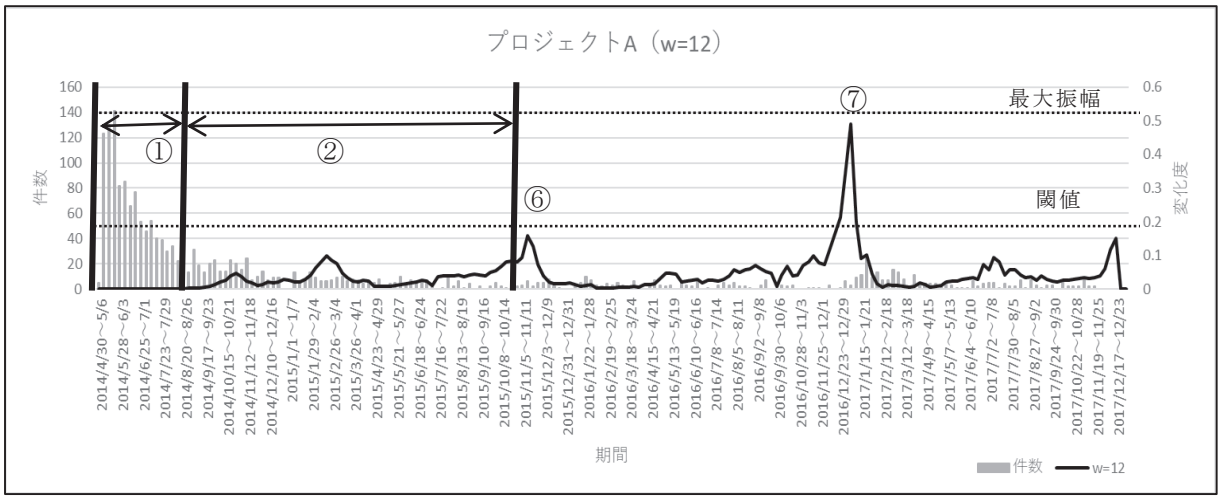

図 19 プロジェクト A，w=12 のグラフ

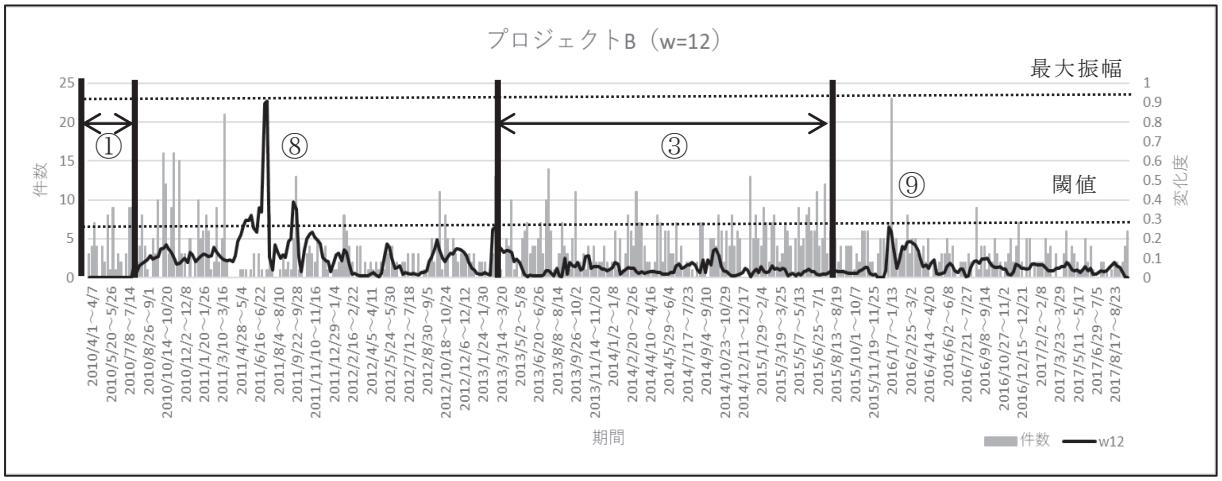

図 20 プロジェクト B，w=12 のグラフ

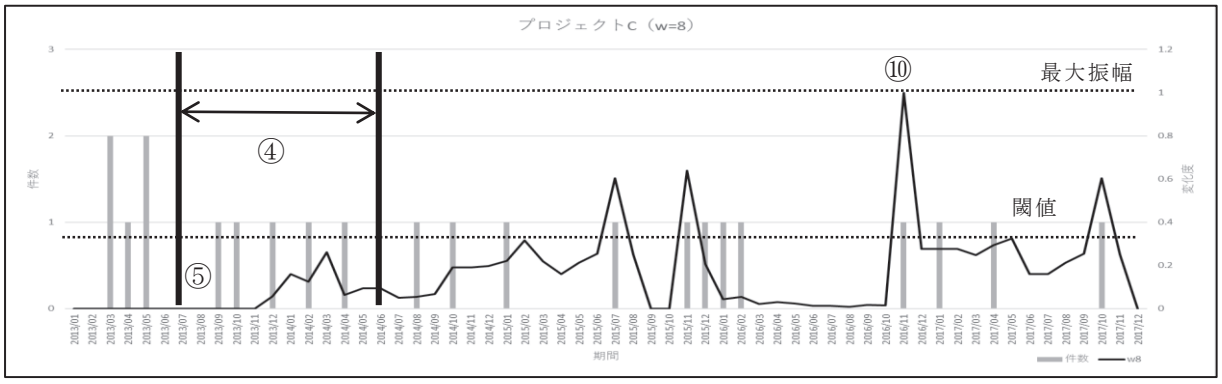

図 21 プロジェクト C．w=8 のグラフ

元データは日次，週次，月次など様々な括り方ができる．できる限り詳細な単位で見た方が変化に早く気づきやすいが，日次データの場合は土日や祝日など，およそ欠陥が検知されにくい日が挟まるため，プロジェクト A，B では週次データを分析対象とした．プロジェクト C は元々の欠陥件数が対象期間に対して少なすぎ，週次だと殆どが 0 件となってしまうことから，月次データを分析対象とした．

(1) 欠陥検知傾向が安定している状況

・ プロジェクトA，Bの①の区間は，データ取得開始後12週間以内であるため，チケットの発行頻度の変動傾向を検知していない．
・ 同プロジェクト②の区間では，チケットの発行は低いレベルを保ち，変化度も同じ傾向を示している．
・ プロジェクトBの③の区間では，四半期の切れ目にチケットが減少する傾向，および期間内は数回程度上下することを繰り返している．
・ プロジェクトCについては，他の2プロジェクトと異なり，月次でのデータの提供であったため，w=12相当での確認ができなかった．一方，w=8において明瞭な傾向が出ていたので，本内容での確認を行うこととした．
・ プロジェクトCの④の区間では，1か月おきにチケットの検知を繰り返している．
・ プロジェクトCの⑤以前の箇所（開始8か月以前）はw=8であるため比較対象外となるので，変化度の検出がない．

グラフ上の上記に示した内容，およびそれに類似した変化度が低い状況を維持している箇所を観察した結果，以下のいずれかを示すものと考えられる．

① 障害チケット検知数の変化がない．
② w値と近いサイクルで障害チケット検知数の周期的な変動がある．
③ 開始後wに相当する期間が経過していない．

(2) 欠陥検知傾向に顕著な変化が発生した状況

それぞれのグラフの変化度が0.2～0.3以上を示している箇所を観察することにより，以下が分かった．

・ プロジェクトAの⑥については，グラフ上判別しにくいが，変化度が0.2を超えている箇所を境に月当りの欠陥検知数がそれまでに比べて4倍程度となっていることが分かった．また当時の担当からヒアリングしたところによると，この時期に大規模なアプリケーションの構造改善を目的としたリファクタリングを実施しており，欠陥の検知量が増えているとの事であった．
・ プロジェクトAの⑦のタイミングでは，欠陥検知数が上昇している．当時のプロジェクトの記録を確認したところ，この時期には大規模な派生開発案件があったが，年度内に開発を完了する必要があるという事情から短期開発となったため，障害が増えていたことを確認した．
・ プロジェクトBの⑧のタイミングでは，およそ12週間前に欠陥検知のピークがあったが，検知時タイミングの欠陥検知数は落ち着いており，それによる変化を検知したものと考えられる．
・ プロジェクトBの⑨のタイミングでは，欠陥検知数の急増に伴い，変化度の上昇を確認した．これについて当時の担当へ確認したところ，外部委託分の障害チケットの一括入力があったとのことで，傾向の変化が表れているということが分かった．
・ プロジェクトCの⑩もしばらく欠陥の検知がなく，9ヶ月ぶりに検知されたことによって変化度の上昇を示していた．このプロジェクトはパッケージベースのプロジェクトでリリース後10年ほど経過していることから，欠陥はほとんど検知されない傾向にある．

グラフ上の上記に示した箇所から，変化度が急増しているケースは以下が共に成立していることを示しているものと考えられる．

① 欠陥検知の急増/急減がある．
② w値と近いサイクルで障害チケット検知数の周期的な変動がある．

なお，顕著な変化を捉えるのに，今回の3プロジェクトでは変化度の最大振幅の30%程度を目安にすると良いようである，ということが分かった．

本分析で得た知見のまとめを付録4に示す．

付録 3. 各種機械学習モデルの比較（井出 剛著「入門 機械学習による異常検知」より筆者らにて整理）

表 1 各種機械学習モデルの比較

機能	入力対象	確率モデル	検出対象	応用	特徴
外れ値検出	多次元ベクトル	独立モデル（単純な閾値）（クラスター外れ値）	外れ値	不正検知 侵入検知	急激な変化に弱い
変化点検出	多次元時系列	近傍法による異常部位検出	時系列上の急激な変化 バースト的異常	ネット攻撃検出 ワーム検出 音声認識	リアルタイム分析
		特異スペクトル解析/変換	時系列上の急激な変化 バースト的異常	心電図による異常検出	ノイズに強い リアルタイム分析
		自己回帰モデル（赤池情報量基準 AIC）	時系列上の急激な変化 バースト的異常	過去データを用いた売上予測	自分自身の過去データからの予測
		線形状態空間モデル（部分空間同定） カルマン・フィルタ（逐次推定法）	測定困難な時系列データの異常・急激な変化	脳圧推定	過去と未来の共通点を探す手法 状態系列推定

付録 4. SSA による可視化に際して,知り得たこと

3 つの実験データの分析において，SSA を行う上で得たノウハウを下記に記す.

① w値の設定値に関して

- パラメータで唯一設定したのはwであるが，最小値はmの 2 倍の値である．また，最大値は元となるデータ期間の長さの 2/3 までである．これらは履歴行列，テスト行列の抽出の仕組み上，その範囲を超えられないためである．（最小値より小さい値，最大値より大きい値を指定するとエラーとなる）
- wで指定した周期に達するまでは，変化度は 0 のままである.

② 変化度を見極めるための工夫

- 変化度を見極める際に，グラフの平滑化を行うと，細かな変化を滑らかにならすことができる（これをスプライン曲線による補間という）．グラフのノイズが除去できるため，似たような変化をより発見しやすくなる.

③ データ量の多寡による変化度の現れ方の違い

- データ量が少ないほど，変化量の振幅が大きくなりやすい.

④ 周期性の読み取り方

- 周期性には 2 つある．1 つは，周期的に同じような欠陥検知の傾向があるケース．この場合，変化度のグラフは振幅が小さくなだらかになる.もう 1 つは，大きな変化が周期的に発生するケース．この場合は似た振幅形状のグラフが複数箇所に現れる.
- 今回の解析で使ったソースコードを用いると，ループ処理でwの値を指定した間隔でしらみつぶしに変化させていくため，一度に複数の変化度のデータが得られる．これを比較しながら，最も変化度の山が際立つものを選び出す.
- あるw値で周期性が得られた場合，その倍数のw値でも同様の波形が得られるが，最初の変化度の山をより早く検知できるように，それらのうち最も小さいw値を選択するのがよい.
- 初期リリースで欠陥が多発し，その後落ち着いてくるプロジェクトでは，最初の変化量の山を過ぎてから，w値の期間が経過したところで履歴行列の欠陥数よりテスト行列の欠陥数が減ることによって，変化量の山が発生する．変化量としては欠陥数が増えても減っても山が立つため，読み取る際に注意が必要.