

Applying Moving Windows to Software Effort Estimation

Third International Symposium on Empirical
Software Engineering and Measurement

Chris Lokan, Emilia Mendes

背景

- 工数見積もりの予測方法としてこれまでのプロジェクトの実績を訓練データとして使用している
 - 全てのプロジェクトのデータを使っており，関連度が低いプロジェクトも含めてしまっている
- 最近のデータで構築した方が適しているのではないか？

リサーチクエスト

- RQ1. 全データを使って構築したモデルと最近のプロジェクトN件に焦点を当てて構築したモデルに精度に違いが出るか？(Nをウィンドウサイズと呼ぶ)
- RQ2.(もし違いがあるならば)ウィンドウサイズが変わることによってどう変わるか調査することにより知見を得られるか

データセット

- ISBSG (International Software Benchmarking Standards Group)のデータから4106件を抽出
- そこからいくつかの条件をもとにデータを剪定
 - 実装日がわからない
 - ISBSGによって高いデータ品質評価 (AまたはB) が割り当てられていない など

実験方法

- 直近N件のデータを使ってモデルの構築
(多変量の逆ステップワイズ回帰 (Harrel's Design package Rのfastbw () 関数) を使用して作成)
- ウィンドウの大きさを変えて実験を行っていく
- 評価基準として
the mean (MMRE) and median (MdMRE) magnitude of relative error (相対誤差) を使用

RQ1.全データを使って構築したモデルと最近のプロジェクトN件に 焦点を当てて構築したモデルに精度に違いが出るか？

Mean absolute residuals with different window sizes

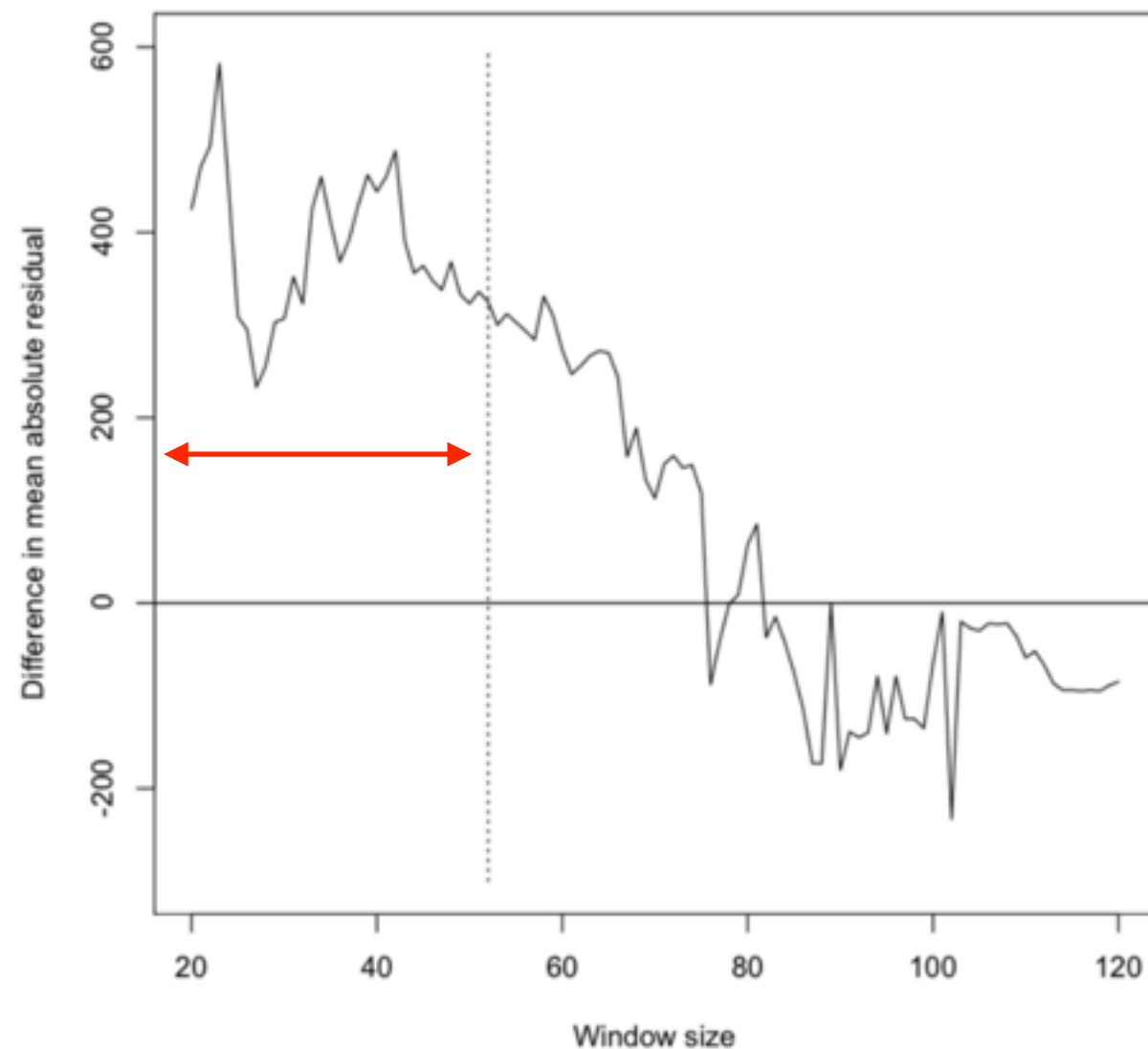
Window size (N)	Testing projects	With window	Without window	p-value
20	201	3137	2712	0.030
30	178	2969	2662	0.035
40	165	3075	2631	0.019
50	153	2947	2624	0.050
60	136	2840	2567	0.104
70	126	2518	2405	0.487
80	126	2468	2405	0.513
90	111	2009	2189	0.154
100	88	2169	2234	0.191
110	75	1762	1821	0.468
120	71	1755	1840	0.113

Mean MRE with different window sizes

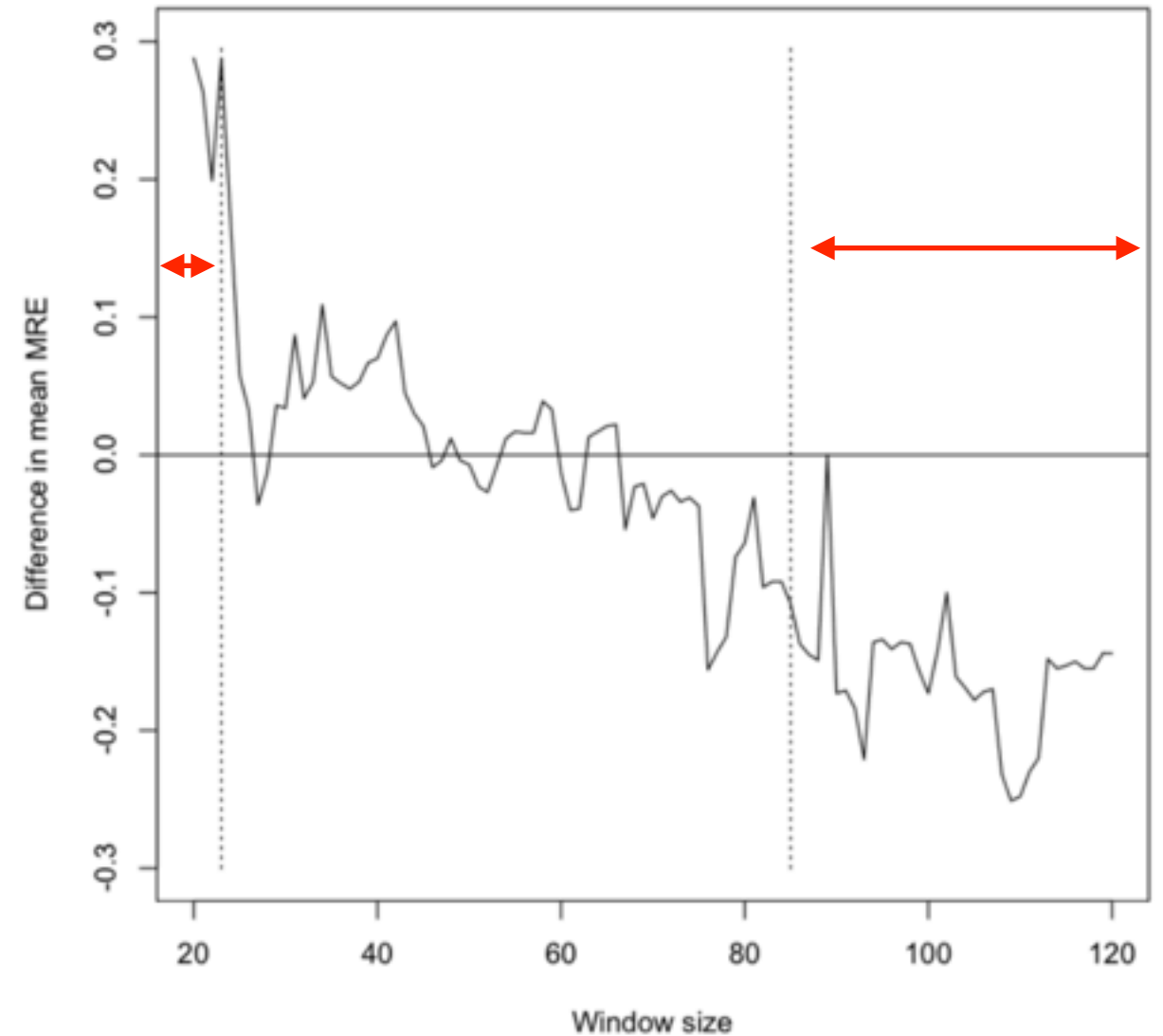
Window size (N)	Testing projects	With window	Without window	p-value
20	201	1.618	1.330	0.029
30	178	1.436	1.402	0.706
40	165	1.483	1.413	0.505
50	153	1.442	1.449	0.942
60	136	1.480	1.493	0.890
70	126	1.491	1.537	0.566
80	126	1.473	1.537	0.416
90	111	1.219	1.392	0.002
100	88	1.232	1.405	0.001
110	75	1.155	1.403	0.011
120	71	1.273	1.417	0.008

RQ1.全データを使って構築したモデルと最近のプロジェクトN件に 焦点を当てて構築したモデルに精度に違いが出るか？

Mean absolute residuals with different window sizes



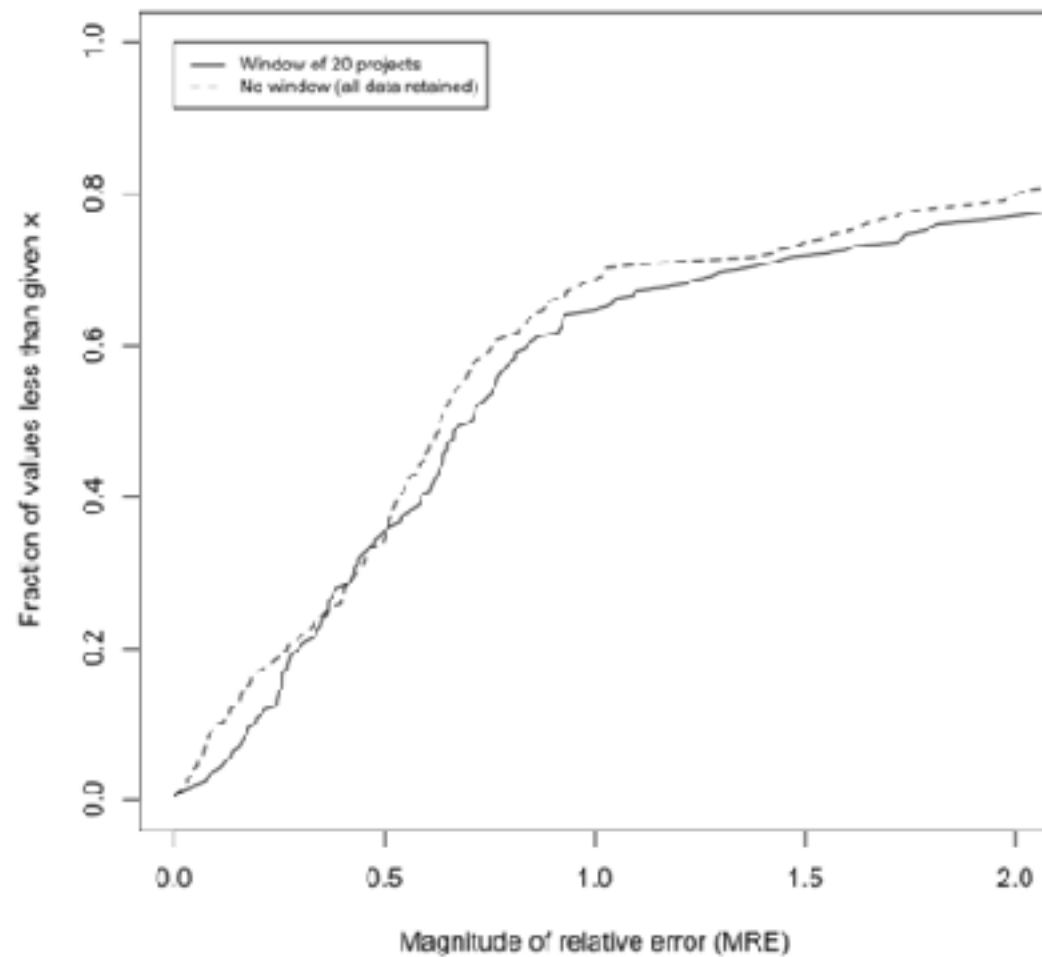
Mean MRE with different window sizes



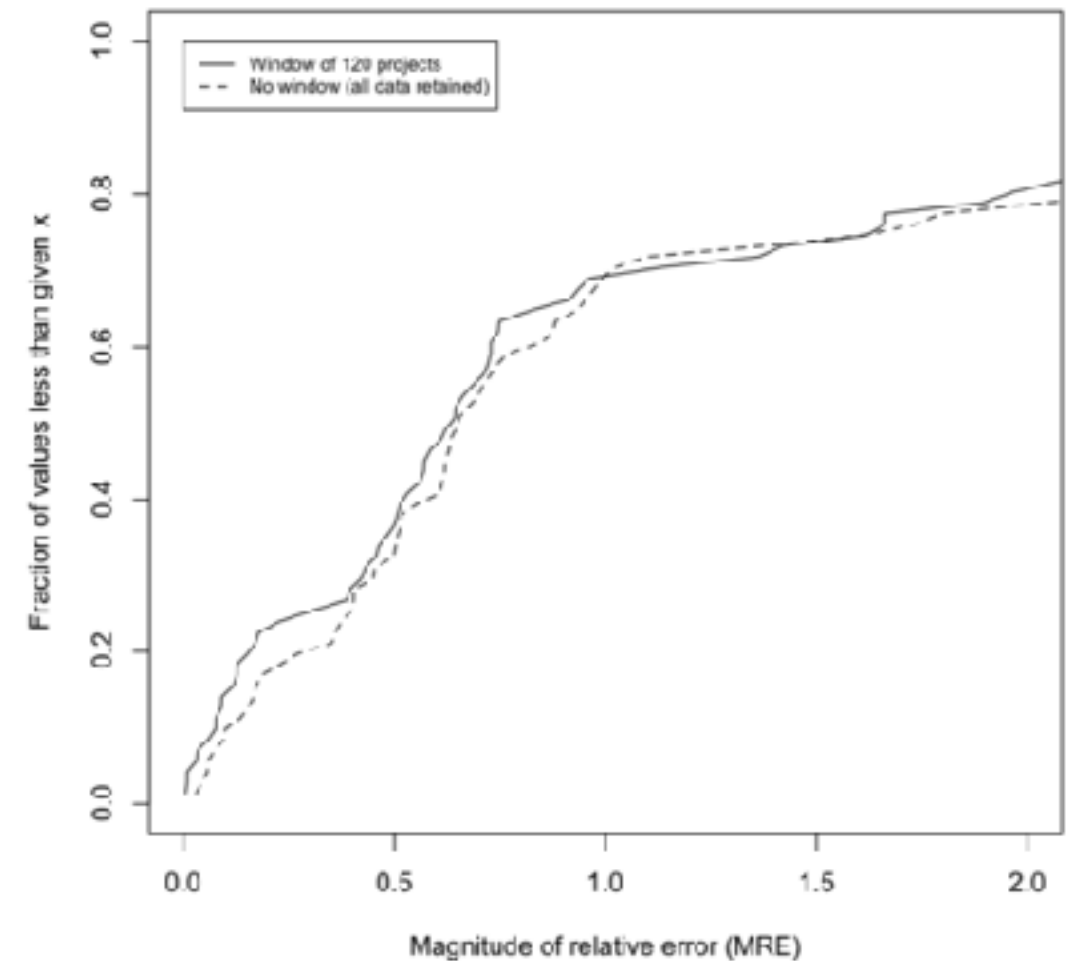
ウィンドウの使用は精度に影響する可能性がある

RQ2.ウィンドウサイズが変わることによってどう変わるか調査することにより知見を得られるか

MMRE, window of 20 projects



MMRE, window of 120 projects



大きなウィンドウを使用すると精度が向上する

結論

- データサイズを変えることにより精度に統計的有意差が出ることがわかった
 - (このデータセットでは)ウィンドウサイズが大きい方がより精度が高くなる
- 別のデータセットではどうなるかはわからない

所感

- 分野は異なるが似た方法の実験を行っているので参考にできそう
- データセットの性質が違うので結果は変わってきそう
(おそらくこの研究の方がトレンドの影響を受けやすい)
- (結論で触れられていたが)固定数でなく期間でやった方が良さそう