

Software Process Extraction and Identification

CS886

Abram Hindle

School Of Computer Science

University of Waterloo

Dec.6 2005

Software Development Process

Requirements, Design

Prototype, Iterations

Patch, Maintenance

Commit Policies

Prototype, Feature Freeze

Bug Fix, Testing, Beta

Mozilla
Apache
Evolution
PostgreSQL
Samba
Linux
*BSD

Software Evolution

Software Evolution

Lehman, Holt,
Godfrey, German

Extract Evidence of Change

SW Process

Extraction

Cooke, Wolf

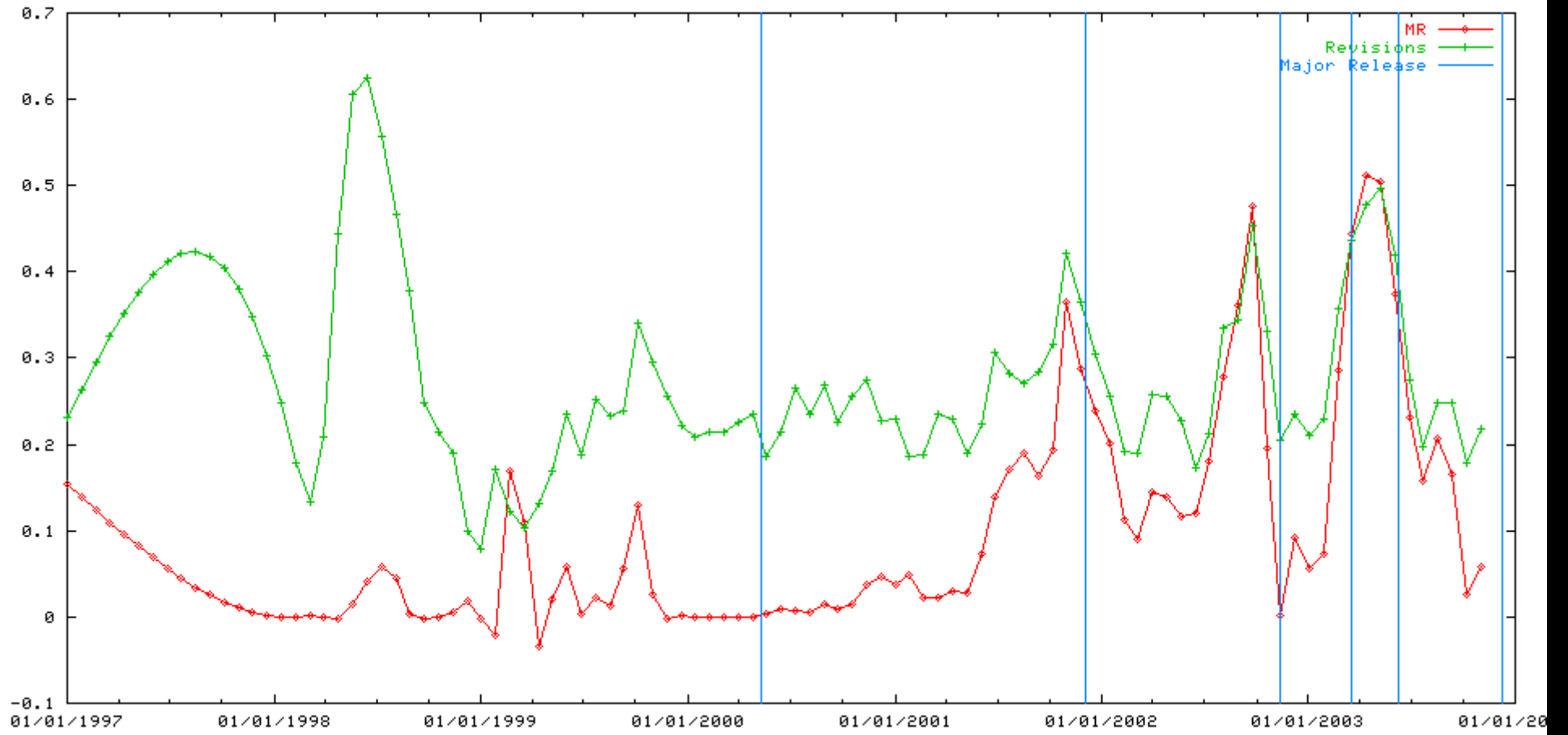
Associate CVS

Behavior with

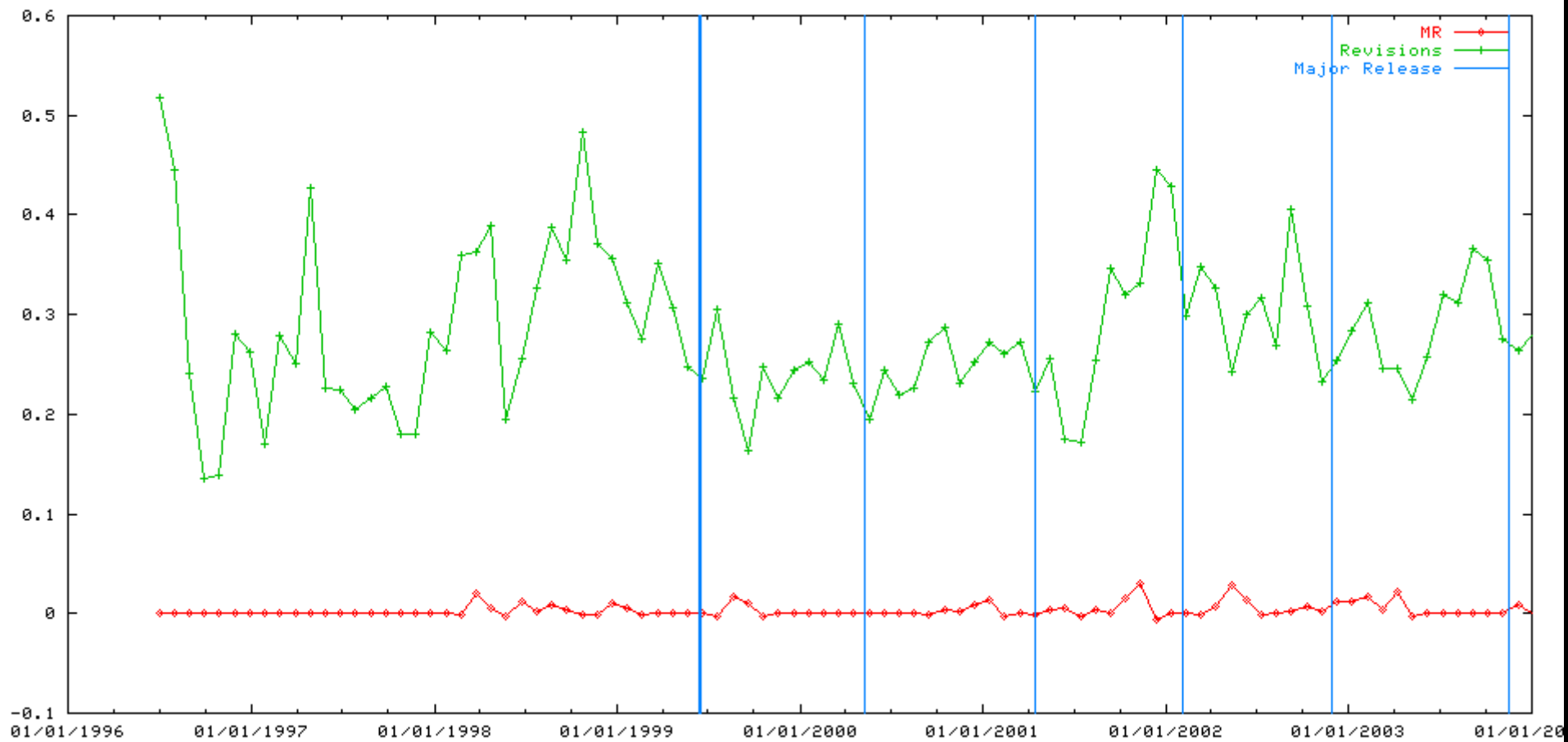
SW Process

**Diff, Patch,
Architecture,
Revisions,
Authors**

Evolution

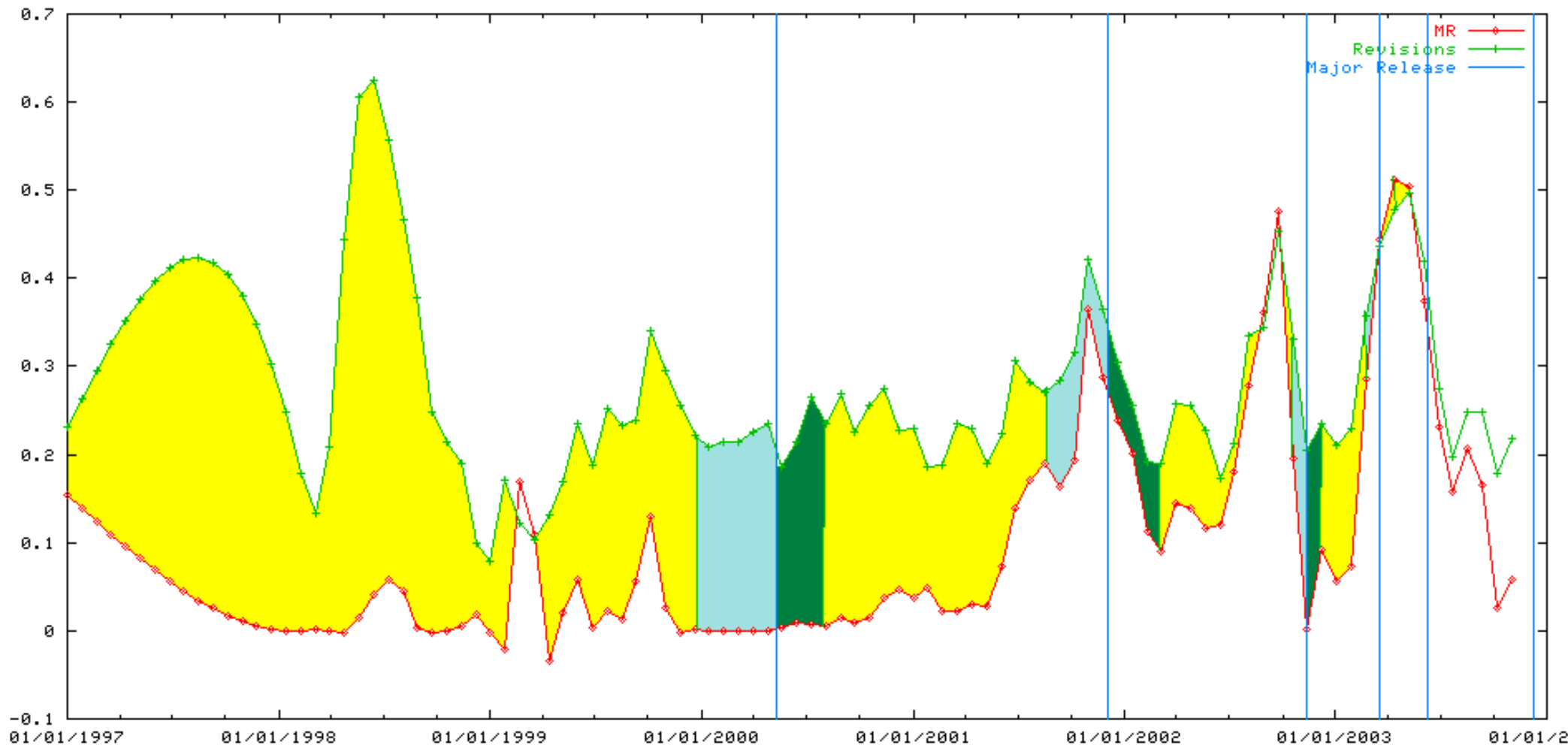


Postgresql



**Train on the
past, classify
the present
(or the future)**

Evolution



Machine Learning

- Noisy

- Poor

- ! Temporal

Markov

Model

- Noisy

- !Temporal

- Discrete

- States++

HMM

HMM

Rabiner, Hollmen,
Tresp

- Fast $O(n^2)$

- Noisy

Hidden State

is hidden label

MDP

- Exponential
- Rewards++
- Learning++

POMDP

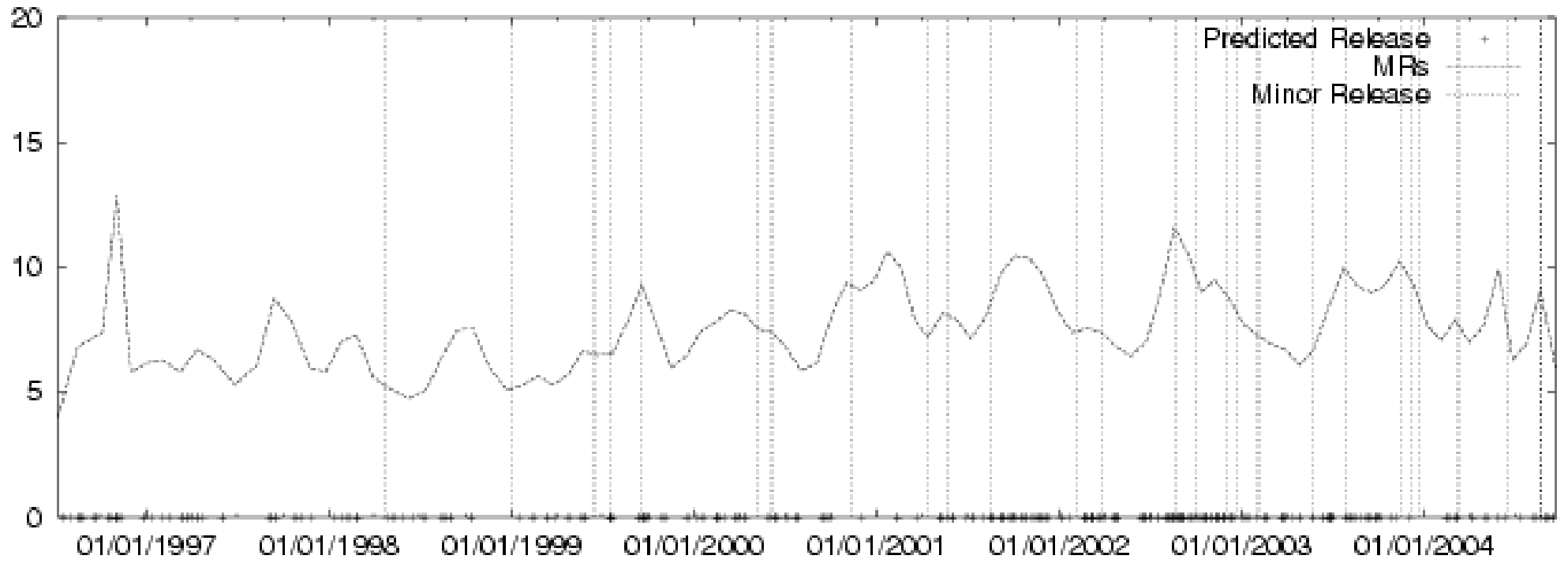
- Expensive
- Retraining
- Continuous

	ML	MM	HMM	MDP	PODMP
Accuracy	-	-	+	+	+
Discrete	+	+	+	+	+
Continuous	+	-	-	+	+
Learning	+	-	+	+	+
Reward	-	-	-	+	+
Speed	+	+	+	-	-
Partial Obs.	-	-	-	-	+
Noisy	+	+	?	?	?
Temporal	-	-	+	+	+

Evolution

PostgreSQL

Postgresql: Predicted Releases versus Minor Releases



Empiricism:

!Markov Model

!Statistical Learning

?HMM

?MDP

?POMDP

POMDP: Accurate Cost++

MDP: Accurate Cost

HMM: Inaccurate Cost--

Questions?