Homicide Flash-up Prediction Algorithm Studying

D.V Serebryakov, I.V. Kuznetsov

NGOGA KIRABO BOB

Outline

1 .Introduction

2 .Data

3 .Numerical experiments

4 .Conclusions

1.Introduction

-The developed algorithm is based on supposition about hierarchicality of crime regime

-The algorithm was specialized for homicide flash-up prediction.

-Compare results of the algorithm realization for its different parameters such as averaging and

accumulation windows and

compare algorithm realization for two different

time series representing two towns, Tambov and Yaroslavl

-*Object* or *object-to-predict* is a time moment where a flash-up of the serious crimes arises plus a special condition

We say that there is *a flash-up* at the time moment *i*, if satisfied

$$R_i = N_i - n_i \geq \sigma,$$

$$n_i = \frac{1}{w_\sigma} \sum_{j=1}^{w_\sigma} N_{i-j}$$

where σ is the intercepting threshold, Ni the number of homicide at the moment i, Wσ the averaging windows and Ri the residue series.

Alarm

there is *a* ***predictor signal*** at the time moment *i*, if satisfied

$$r_i \equiv b_i - k_i \geq \beta,$$
$$k_i = \frac{1}{w_\beta} \sum_{j=1}^{w_\beta} b_{i-j} \ ,$$

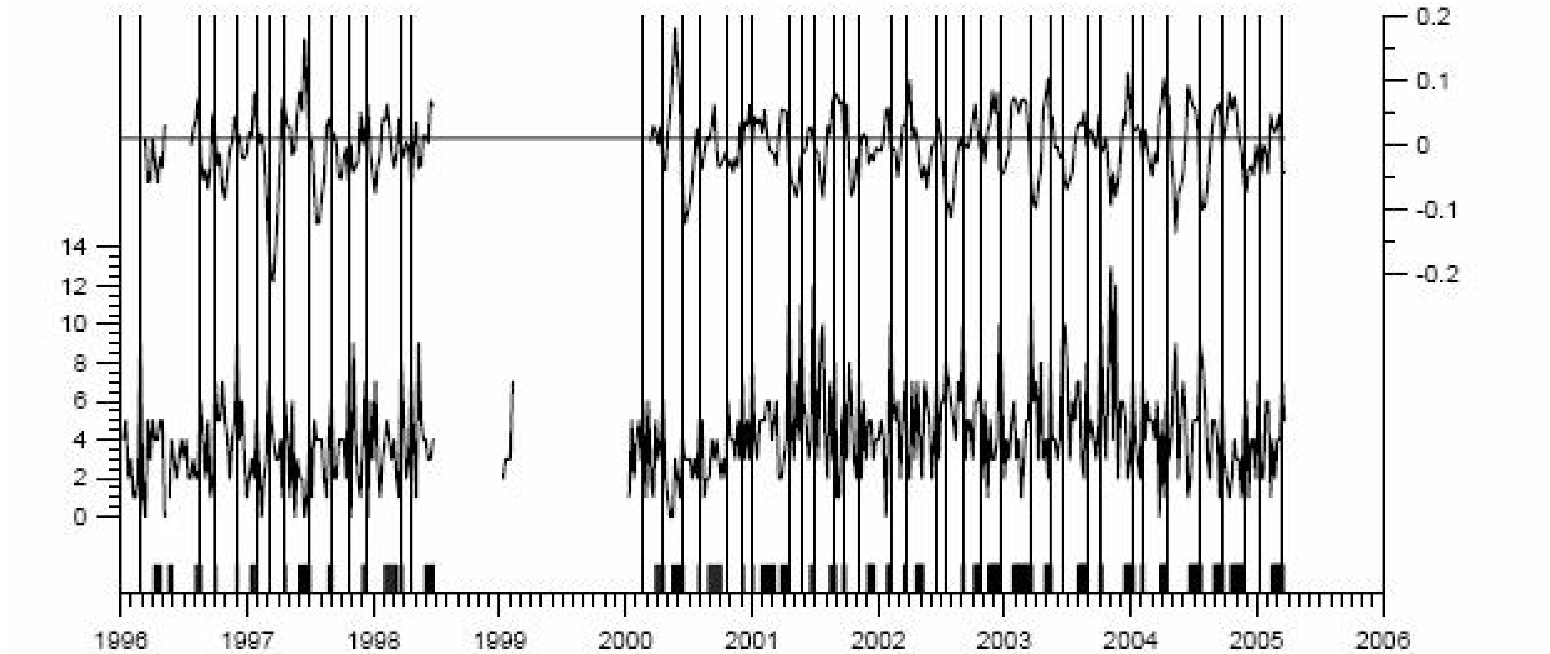
where β is the intercepting threshold, bi=(*V1*(*i*) – *V3*(*i*)) / 2, Wβ is the averaging window and ri is the residue series.

If a predictor signal is presents at time moment i, then one declares an alarm for following d serial moments i+1, i+2,…, i+d.

If there is an object **s** during the alarm, then the final alarm moment is **s** after which this alarm is cancelled. If an object **s** and a predictor signal **i** present at the same moment or 0 ≤ **i – s** ≤ 2, then an alarm is not declared

- *an* ***alarm is successful*** if the alarm consists an object-to-predict, and *an alarm is false* (***false alarm***) if there is no object over the alarm interval.
- *the* ***object is predicted*** if it is in an alarm, and an object-to-predict is ***fail to predict*** (***a fail-to-predict object***) if it is not covered any alarm interval.

A result of the algorithm realization



Investigation of some relations between algorithm parameters and its results is a goal in this work.

2.Data

weekly crime data from towns Tambov and Yaroslavl.

Monitoring time for Tambov is period from 10.01.1996 till 23.03.2005

without period from 01.07.1998 till 05.01.2000,

for Yaroslavl period is from 9.02.1993 till 19.06.2001,

The beginning of weeks is defined as Tuesday.

The monitoring time periods are 405 and 436 weeks, i.e. there are 405 and 436 cases or observations in initial time series for Tambov and Yaroslavl correspondently.

3.Numerical Experiments

- Here they are varying for every town windows $w\beta$ and w, σ and β, when $w\sigma = 5$.
- They consider values of $w\beta$ and w = 5, 7, 10, 17 and 26 weeks,

σ and β are located in intervals [1; σmax] with step $\delta\sigma = 0,1$
and [0,001; βmax] with step $\delta\beta = 0,001$ correspondently.
For every algorithm realization they record the following information:

- 1) N (number of objects),
- 2) N– (number of fail-to-predict objects),
- 3) Ta (duration of alarms),
- 4) W (number of alarms),
- 5) W– (number of false alarms)

The quality of prediction is estimated by the set of quantities
$\eta = N^- / N$, $\tau = T_a / T$, $\varphi = \eta + \tau$, $\chi = W^- / W$ and

$$\psi = \chi + \varphi$$

If $\eta=1$ then $\tau=0$,all objects are fail to predict and if $\eta=0$ then $\tau=1$,they predict all objects

An integral quality of prediction is evaluated by the quantity $\varphi = \eta+\tau$.

Condition $\varphi < 1$ corresponds to a non-trivial prediction.

The smaller is φ the better is the prediction.

Considering values of η and τ:

Non-trivial prediction is considered as **acceptable** if $\eta < 1/2$ and $\tau < 1/2$ and **successful** if $\eta < 1/3$ and $\tau < 1/3$.

Minimal values of quantities η, τ, χ, φ, ψ for fixed w_β and w.

town	w_β	w	η	τ	χ	φ	ψ
Tambov	5	5	,000	,014	,000	,322	,529
	5	7	,000	,013	,000	,169	,518
	5	10	,000	,023	,000	,302	,583
	5	17	,000	,022	,000	,243	,614
	5	26	,000	,013	,000	,235	,664
	7	5	,000	,013	,000	,419	,518

Tambov	7	7	,000	,016	,000	,155	,544
	7	10	,000	,024	,000	,270	,565
	7	17	,000	,022	,000	,198	,621
	7	26	,250	,014	,000	,443	,693
	10	5	,000	,023	,000	,552	,583
	10	7	,000	,024	,000	,520	,565
	10	10	,000	,010	,000	,273	,595
	10	17	,000	,014	,000	,184	,666
	10	26	,250	,011	,000	,503	,681
	17	5	,000	,022	,000	,380	,614
	17	7	,000	,022	,000	,373	,621
	17	10	,000	,014	,000	,434	,666
	17	17	,227	,018	,000	,519	,707
	17	26	,238	,012	,000	,498	,692
	26	5	,000	,013	,000	,291	,664
	26	7	,250	,014	,000	,533	,693
	26	10	,250	,011	,000	,620	,681
	26	17	,238	,012	,000	,665	,692
	26	26	,333	,008	,000	,528	,759
Yaroslavl	5	5	,111	,016	,000	,432	,556
	5	7	,000	,019	,000	,528	,573
	5	10	,000	,020	,024	,376	,627
	5	17	,000	,019	,026	,498	,692
	5	26	,083	,021	,000	,451	,564
	7	5	,111	,019	,000	,523	,523
	7	7	,000	,021	,000	,418	,594
	7	10	,000	,018	,000	,507	,573
	7	17	,000	,017	,000	,521	,607
	7	26	,000	,019	,000	,538	,561
	10	5	,000	,030	,000	,527	,542
	10	7	,000	,020	,000	,347	,603
	10	10	,000	,020	,000	,509	,572
	10	17	,000	,017	,000	,423	,605
	10	26	,000	,021	,000	,531	,586
	17	5	,000	,022	,000	,537	,586
	17	7	,000	,025	,000	,439	,633
	17	10	,000	,023	,000	,583	,583
	17	17	,000	,021	,000	,538	,564
	17	26	,000	,022	,000	,570	,591
	26	5	,000	,022	,000	,500	,559
	26	7	,000	,016	,000	,324	,558
	26	10	,000	,016	,000	,418	,548
	26	17	,000	,017	,000	,340	,578
	26	26	,000	,016	,000	,425	,575

-percentage of false alarms, χ.

Prediction is considered as effective if $\chi < 1/2$.

-To evaluate together these three quantities

they consider the sum $\psi = \eta + \tau + \chi$, which

values from limitations above for appropriate

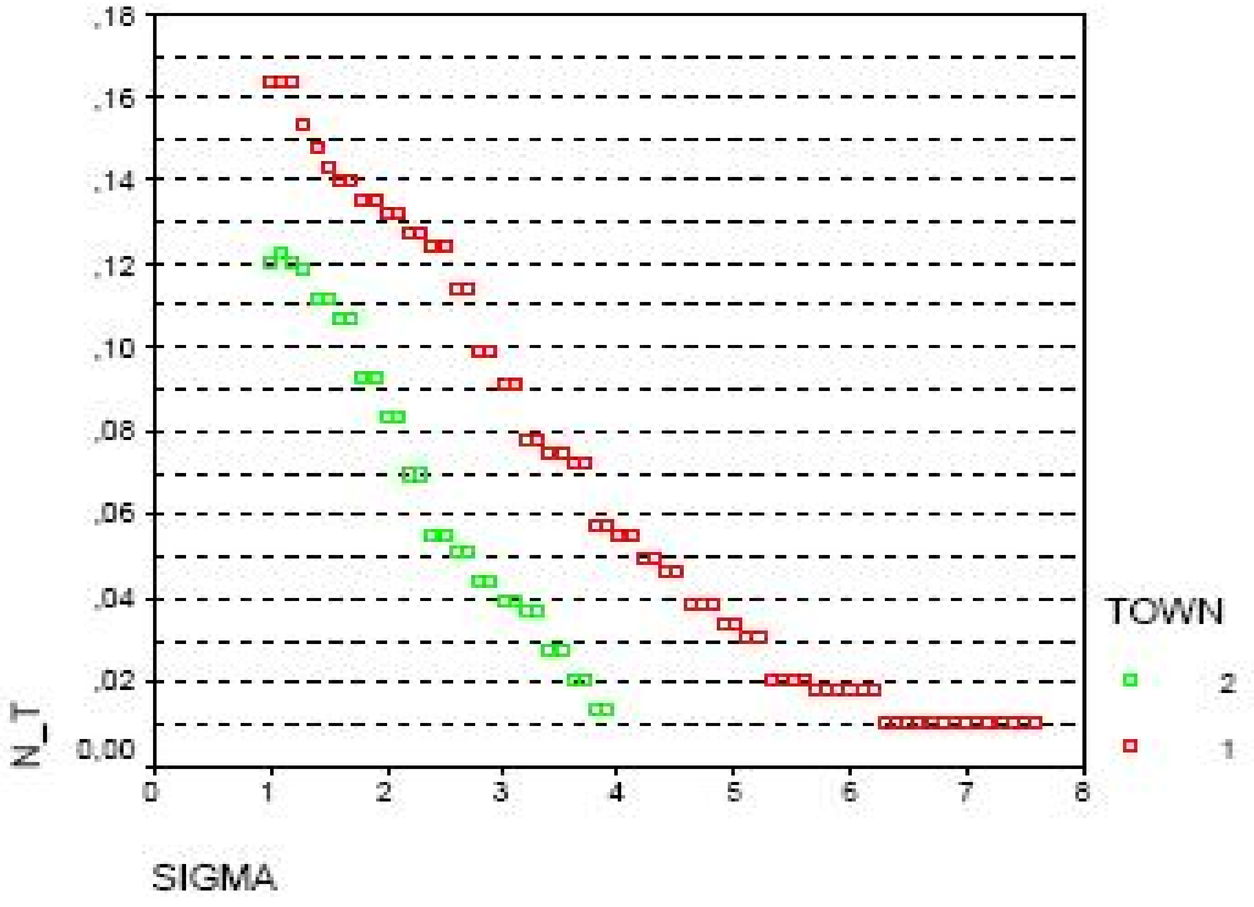
prediction should be less 1,5.

-to distinguish objects by its frequency they introduce the ratio of number of objects N to monitoring time T for corresponding town,

$n = N/T$.

n depends on σ and the higher is σ the smaller is n

Dependence n(σ).



The ratio N/T is considered as the power of event,

power of flash-up which is denoted by **p**.

- According to the form of dependence of

$n(\sigma)$ there are four intervals of n:

$p = 4$ if $n \leq 0,03$,

$p = 3$ if $0,03 < n \leq 0,06$,

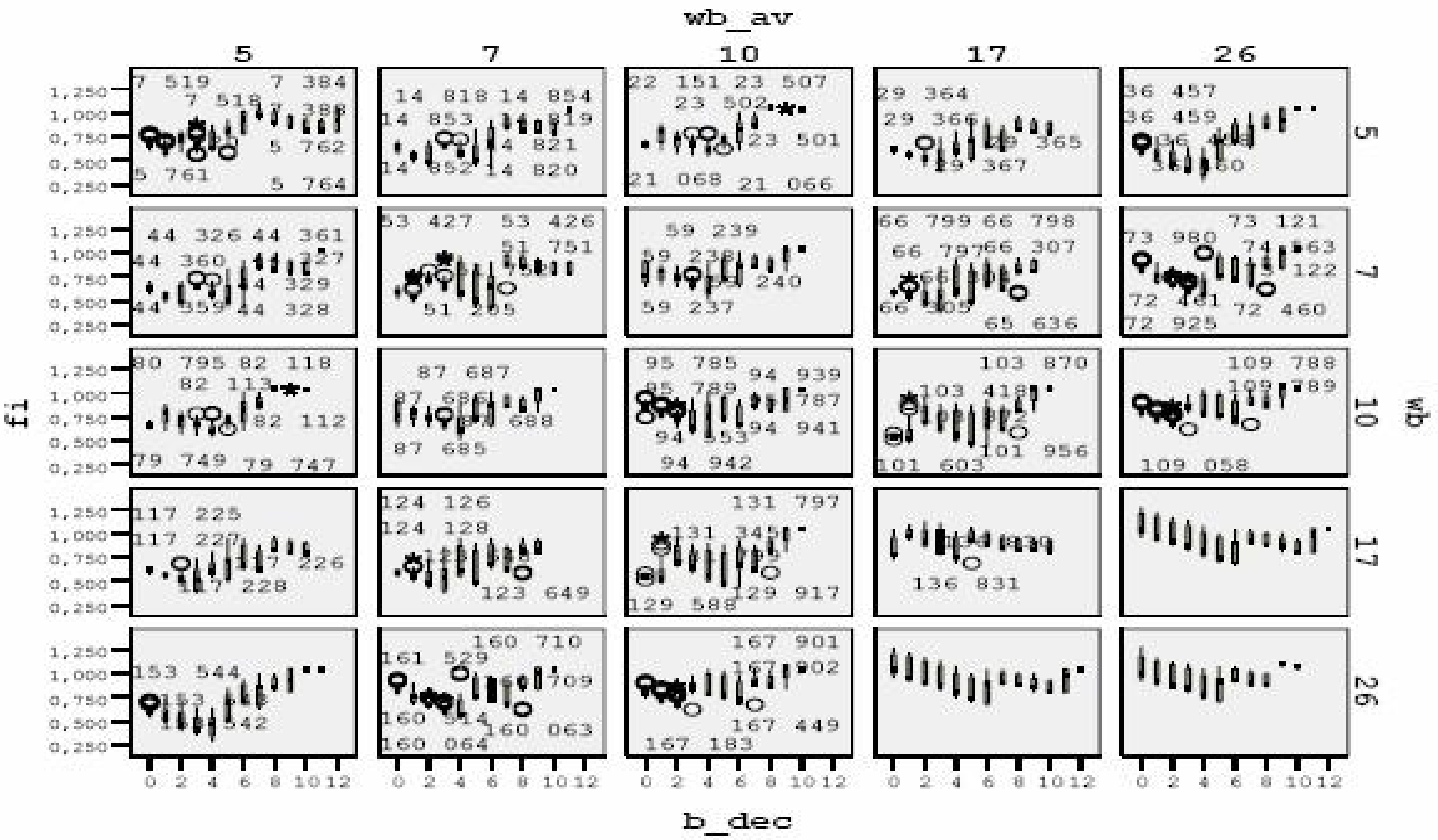
$p = 2$ if $0,06 < n \leq 0,115$ and

$p = 1$ if $n > 0,115$.

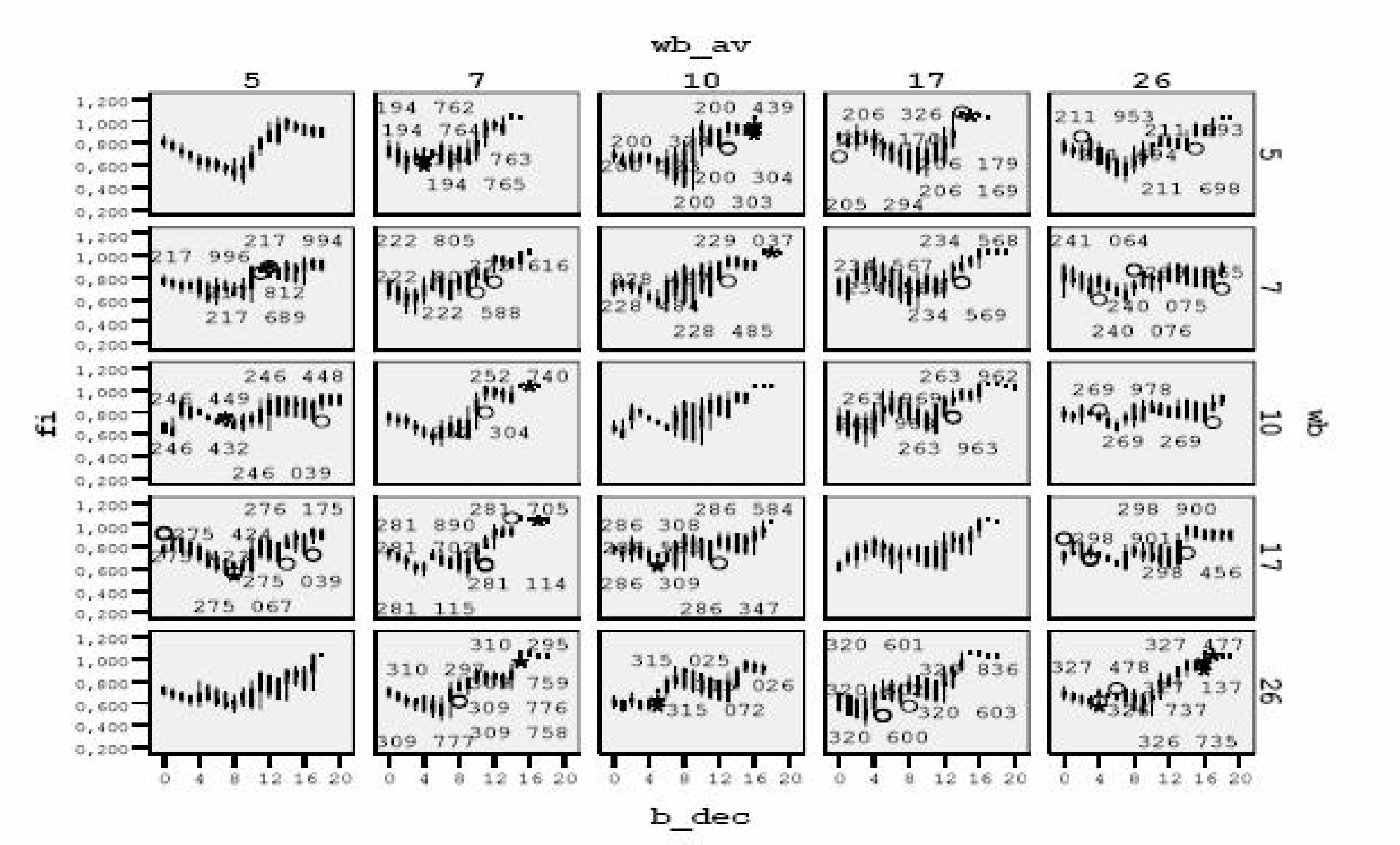
Minimal and maximal values of quantities η, τ, χ, φ, ψ for fixed p.

town	p	η		τ		χ		$\varphi = \eta + \tau$		$\psi = \eta + \tau + \chi$	
		min	max	min	max	min	max	min	max	min	max
Tambov	1	,20	,98	,01	,43	,00	,50	,52	**1,01**	,52	**1,51**
	2	,11	1,00	,02	,58	,00	1,00	,62	**1,05**	,76	**2,05**
	3	,05	1,00	,03	,65	,24	1,00	,50	**1,08**	,90	**2,08**
	4	,00	1,00	,03	,71	,70	1,00	,16	**1,27**	,96	**2,16**
Yaroslavl	1	,15	,96	,02	,46	,00	,57	,52	**,99**	,52	**1,60**
	2	,08	,96	,02	,60	,00	,67	,53	**,99**	,55	**1,65**
	3	,05	,96	,02	,66	,17	,83	,56	**1,01**	,90	**1,83**
	4	,00	1,00	,02	,70	,55	1,00	,32	**1,08**	1,08	**2,08**

Dependence φ (Δβ) for Tambov



Dependence φ (Δβ) for Yaroslavl



4.Conclusions

- There are groups of data where small variations of algorithm parameters lead to small variation of prediction results.
- The algorithm represents a simular result over times series for the 2 different towns.
- The algorithm is good enough and has a property of universality.
- The following step in studying the algorithm is to decrease the percentage of false alarms for rare events.

MURAKOZE