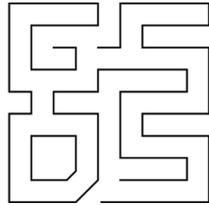
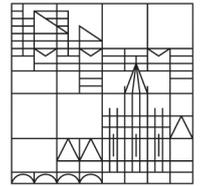


GRADUATE SCHOOL  
OF DECISION SCIENCES



Universität  
Konstanz



**GSDS**  
**Working Paper**  
**No. 2021-01**

# Real-Time Change Point Detection of the COVID-19 infections

**Lukas Ringlage**

July 2021

## **Graduate School of Decision Sciences**

All processes within our society are based on decisions – whether they are individual or collective decisions. Understanding how these decisions are made will provide the tools with which we can address the root causes of social science issues.

The GSDS offers an open and communicative academic environment for doctoral researchers who deal with issues of decision making and their application to important social science problems. It combines the perspectives of the various social science disciplines for a comprehensive understanding of human decision behavior and its economic and political consequences.

The GSDS primarily focuses on economics, political science and psychology, but also encompasses the complementary disciplines computer science, sociology and statistics. The GSDS is structured around four interdisciplinary research areas: (A) Behavioural Decision Making, (B) Intertemporal Choice and Markets, (C) Political Decisions and Institutions and (D) Information Processing and Statistical Analysis.

GSDS – Graduate School of Decision Sciences  
University of Konstanz  
Box 146  
78457 Konstanz

**Phone:** +49 (0)7531 88 3633

**Fax:** +49 (0)7531 88 5193

**E-mail:** [gsds.office@uni-konstanz.de](mailto:gsds.office@uni-konstanz.de)

**-gsds.uni-konstanz.de**

**ISSN: 2365-4120**

July 2021

© 2021 by the author(s)

# Real-Time Change Point Detection of the COVID-19 infections\*

Lukas Ringlage

University of Konstanz<sup>†</sup>

This version: June 25, 2021

## Abstract

The dynamics of the COVID-19 infections is characterized by many fundamental change points due to a quickly changing environment (e.g. policy adaptations, changing behaviour of the population or mutations of the virus). In a framework of numerous time-varying influencing factors, popular approaches from the disciplines of predictive or causal modelling suffer from unrealistic model assumptions and limited validity of the model results. As an alternative approach of online change point detection (CPD) a decision model based on a real-time *alarm statistic* is proposed that can be tuned by an adaptive objective function forcing the early identification of change points in the pandemic. In this model, one key aspect is the analysis of micro-level information, which can be exploited to gain a more fundamental picture of the pandemic development. For the case of Germany, the utility of micro-level information for improving the model performance is demonstrated. The ultimate goal of the CPD model is to anticipate periods of severe infection waves and to recognize when pandemic policies take effect. Hence, the proposed model might serve as a valuable early-alert system for policymakers to support timely and effective measures.

**Keywords:** Online Change Point Detection, Panel Data Heterogeneity, Aggregation, Corona pandemic

**JEL classification:** C32, C51

---

\*I would like to thank Winfried Pohlmeier and the participants of the Econometrics Colloquium (University of Konstanz) for the helpful comments and discussions. Financial support by GSDS is gratefully acknowledged, Declarations of interest: none

<sup>†</sup>Department of Economics, Universitätsstraße 10, D-78457 Konstanz, Germany. Phone: +49-7531-88-5373, Email:[Lukas.Ringlage@uni-konstanz.de](mailto:Lukas.Ringlage@uni-konstanz.de).

# 1 Introduction

In the current Corona pandemic, the development of infections is characterized by a frequently changing environment. There exist numerous time-varying key drivers influencing the infection dynamics that exacerbate the prediction of infections or the identification of causal effects due to structural changes (Hyndman, 2020). As depicted in Figures 5(a), 5(b) and 5(c), the most popular measure of the infections, the so-called seven-days-incidence<sup>1</sup>, frequently changes unexpectedly at the beginning, during and at the end of infection waves (especially in periods March 2020 - May 2020 and September 2020 - January 2021). A priori, we do not know when the next structural change occurs and how the change will look like. In view of these frequent structural changes in the environment, models detecting causal inference clearly would have to account for temporal heterogeneity which comes along with a large set of restrictive assumptions (Bach et al., 2020) to be able to pin down selected causalities. Moreover, long-horizon predictions of the infections based on previous data are not credible in view of frequently changing model parameters (Cássaro and Pires, 2020). Finally, many other studies (e.g. Dong et al. (2020)) analysing how different countries mastered the first infection-wave suffer from limited generalizability and are problematic due to a low degree of comparability of reporting routines across countries.

Taking these difficulties into account, my approach to the problem of predicting the future development of the Corona pandemic is the identification of substantial changes in the development of COVID-19 infections rather on finding the best predictive model. In the previous months, according to Figure A.5.2, change points have occurred suddenly and have been milestones to a new period of changed infection dynamics. Put differently, change points are a first indicator for new phases of persistently higher or lower infection numbers. Hence, change points in the pandemic development have a very high importance for policymakers who might wish to adapt pandemic measures at an early stage. As presented by many studies (e.g. Dave et al. (2021), Chen et al. (2020), Haug et al. (2020)), the timing of introduced measures is decisive for the efficacy of political measures. Consequently, there is need for an alert system which identifies change points as early and accurate as possible.

The core idea of the identification of change points, also known as the task of *change point*

---

<sup>1</sup>New infections during the last seven days per 1 million persons

*detection (CPD)* is to cut a given time series into homogeneous segments w.r.t. certain properties or parameters of the underlying model. In order to develop a CPD model there is need for a clear-cut definition of what a change point is.

When analysing change points in the COVID-19 infections, the lack of a clear-cut definition for an epidemiological change point necessitates a more detailed discussion of the term 'change point'. Eventually, there exist neither *true* change point dates nor a consensus about the change point dates in Germany. One reason for that might be the heterogeneity of the infection dynamics across different sub-regions. Particularly, the problem in this context is that there is no clear threshold for change points defining the required magnitude of a change to report a change point. The lack of commonly accepted change points makes the comparison of the alert dates between the proposed models more complicated. There exists a wide range of definitions of the terminus 'change point' depending on the application and the model which is used to identify these *change points*. In the framework of a regression model

$$y_t = x_t \beta_t + \varepsilon_t, x_t \in \mathbb{R}^{1 \times k}, \beta_t \in \mathbb{R}^{k \times 1}, \quad (1)$$

changes in the dependent variable  $y$  as the infection incidence might be deduced to changes in the regressors  $x_t$ , e.g. due to changing behaviour of the population, or in the parameters  $\beta_t$ , e.g. due to mutations of the virus. According to the definition of the seminal work of [Basseville et al. \(1993\)](#) in the field of CPD, a change point depicts an abrupt transition between two states subject to 'characteristic properties or parameters' ([Basseville et al. \(1993\)](#), p. IX) of the underlying data generating process of a time series. The goal of this work is to develop a model to detect change points following this definition.

The task of CPD can be approached in two different designs that differ in the perspective of the analysis of a time series. At the one hand, from a retrospective view, the focus is on finding potential change points in historic data in the fixed interval  $t = 1, \dots, T_0$  (*offline* CPD). On the other hand, *online* CPD aims at detecting change points in current, regularly updated data at time  $t = T_0 + 1, \dots, T_0 + \tau$ . Hereby, the goal is to monitor changes in  $\beta_t$  of model equation (4) in real-time to anticipate e.g. a new wave of infections. Hence, *online* CPD has a particularly high relevance in the current Corona pandemic as political decisions could be derived from a data-driven CPD model in real-time. Eventually, the ultimate goal for the CPD model is to anticipate serious in-

fection waves as early as possible in the current COVID-19 pandemic for the example of Germany.

There are many model types based on various statistical concepts that encounter the task of CPD comprising density ratio-estimation (Kawahara and Sugiyama, 2012), residual analysis techniques consisting of CUSUM and MOSUM methods (Zeileis, 2003), bootstrapping (Buzun and Avanesov, 2017) or Fused LASSO estimation (Bleakley and Vert, 2011; Alaíz et al., 2013; Harchaoui and Lévy-Leduc, 2010) etc. Jiang et al. (2020) attempt to detect CPs based on a sequential change point test based on a hybrid of the SN (nascent self-normalization) method (Shao, 2010) and the NOT (Narrowest-Over-Threshold) algorithm (Baranowski et al., 2016) (offline CPD). Moreover, the general task of detecting common change points in heterogeneous panel data has been addressed in Okui and Wang (2020) and Lumsdaine et al. (2020) imposing a partial linear model structure with potentially common parameters over the temporal and cross-sectional dimension. Hereby, the authors allow for structural changes in both panel dimensions, i.e. if a group membership changes or the parameter vector of a group changes.

In view of the enormous variety of techniques, the mentioned models strongly differ in its structural assumptions and in the requirement, whether the number of CPs must be known ex ante. In the considered application of COVID-19 pandemic, the number of CPs is by assumption unknown, such that I employ a model that generates the number of change points as an outcome. For that reason, I develop a Fused LASSO model which chooses the optimal number of CPs by means of total variation regularization (Truong et al., 2020).

In the framework of time-varying parametric models, the Fused LASSO method can be applied to identify time intervals with constant dynamics in the COVID-19 infections. In more detail, the main idea of Fused LASSO is to penalize the total variation in the parameters with the goal to stabilize the parameter vector over time. By design, the Fused LASSO model itself detects change points only in the retrospective view. Thus, the main contribution to the existing literature in the field of CPD is the extension of the Fused LASSO estimation to an online approach monitoring change points in real-time.

The analysis of the dynamics of a pandemic within a country can be approached at different

levels of granularity. Less surprisingly, the information value might substantially depend on the granularity level of the analysis. Clearly, the dynamics of the Corona pandemic in a country is characterized by many different trends in sub-regions. This heterogeneity becomes visible when examining the infection numbers at a more detailed level. Figure A.5.2 depicts the incidence of the country (macro-level), selected federal states (mezzo-level) and selected counties (micro-level) for three different periods. A more detailed analysis of the heterogeneity is provided in A.5.2. As a consequence, taking micro-level information into account yields a much better understanding of the macro-level data. My contribution to this field is to derive a (set of) macro-level measures, later called *alarm statistics*, that incorporate all relevant CP information including the degree of heterogeneity on the micro-level.

I make use of real-time CPD concepts stemming from the statistically related field of business cycle dating that are designed to detect change points in a national business cycle determined by a group of heterogeneously behaving micro-level entities (such as companies or business sectors). Stock and Watson (2010) categorize the business cycle dating models into *Date-Then-Aggregate (DTA)* and *Aggregate-Then-Date (ATD)* approaches.

Based on an aggregated alarm statistic that captures changed model dynamics based on residuals of the AGFL model, I will employ an ATD-type decision model which monitors the incidence dynamics on the country-level. Specifically, similar to the idea in Grillenzoni (2012), the hyper-parameters of the decision model will be tuned by an objective function tailored to the given application. Tuning a CPD model w.r.t. an objective function is an innovative approach in general, particularly to the case of detecting CPs in the Corona pandemic. Though, it is an illustrative method to detect CPs. First, this objective function is a mathematical definition of the quality attributes of the CPD model and secondly, depicts an intuitive basement for tuning the model hyper-parameters determining the sensitivity of the CPD model. Based on the limited explanatory power of ATD alarm statistics, I will design supplementary DTA measures which can be used to refine the expressiveness of the change points detected based on the ATD model.

This paper is organized as follows. In Section 2, I will provide a theoretical overview over the total variation regularization (TVR) techniques used in the given application case and elaborate

on real-time CPD systems which are based on the TVR methods. Hereby, the focus will be on the derivation of useful alarm statistics for monitoring the changes in the dynamics of the infections. Moreover, I elaborate on an objective function used as the fundament for tuning the online CPD model. In Section 3, the data set at hand is described and the CPD model is proposed. The results are reported in Section 4. Subsequently, challenges are presented in Section 5. Furthermore, some limitations of the given approach and further ideas are discussed.

## 2 The Change Point Detection Model

### 2.1 Micro-, Mezzo- & Macro-level

The application of this paper is to detect change points in the Corona pandemic in Germany as early as possible. The presented models operate on three different granularity levels of the data. The macro-level applies to German-wide data and the micro-level to county-level data (*Landkreis* resp. *Stadtkreis*). In some parts of the analysis, German regions (mezzo-level) will be formed in order to demonstrate the heterogeneity in macro-level measures and in order to augment the testing examples for evaluating the model performance. As the incidence  $y_{i,t}$  is relative to the population size of the respective county, the aggregation of the micro-level data to the macro-level accounts for the population proportions of the single counties relative to the whole population

$$y_t = \sum_{i=1}^N y_{i,t} \cdot \underbrace{\frac{pop_i}{\sum_{j=1}^N pop_j}}_{=\omega_i}. \quad (2)$$

In the following, the operation can be used to aggregate not only the dependent variable or equivalently, the residuals  $e_t = y_t - \hat{y}_t$ , but also alarm statistics and further measures proposed in the subsequent section. Analogously to (2), the aggregate for region  $R_m$  is calculated by

$$y_{m,t} = \sum_{i \in R_m} y_{i,t} \cdot \underbrace{\frac{pop_i}{\sum_{j \in R_m} pop_j}}_{=\omega_i}, m = 1, \dots, M. \quad (3)$$

### 2.2 Offline CPD

Facing a panel of German counties on the incidence of COVID-19 infections, the data exhibit a cross-sectional and a temporal dimension. Initially, the cross-sectional dimension is dropped by considering a single county over time  $t = 1, \dots, T$ . Assume a linear model of the form

$$y_t = x_t \beta_t + \varepsilon_t, t = 1, \dots, T \quad (4)$$

with time-varying parameter vector  $\beta_t \in \mathbb{R}^K$ . Estimating the flexible, time-varying model (4) without regularization clearly overfits the true model. In this context, penalties on parameter

changes are introduced to smooth parameter changes. A popular regularization technique is the Fused LASSO (Tibshirani et al., 2005). Applied on time series data, the purpose of the LASSO estimation will be to segment the model into periods of homogeneous model parameters  $\beta_t$ .

For a  $K$ -dimensional parameter vector  $\beta_t$ , the AGFL (Qian and Su, 2016) is used to estimate the parameter matrix  $\beta$

$$\min_{\beta \in \mathbb{R}^{T \times K}} \frac{1}{2} \|X\beta - y\|^2 + \lambda_1 \|\beta\|_{2,1} + \lambda_2 \|\dot{w} D_2 \beta\|_{2,1} \quad (5)$$

$$\text{with } \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_T \end{pmatrix}_{[T \times K]}, \quad X = \begin{pmatrix} x'_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & x'_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & x'_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 & x'_T \end{pmatrix}_{[T \times TK]} .$$

In this equation, the first term aims to optimize the model fit by minimizing the residuals  $e = X\hat{\beta} - y$ . Moreover, the group total variation (GTV)

$$GTV(\beta) = \|\beta\|_{2,1} = \sum_{t=1}^T \|\beta_t\|_2 = \sum_{t=1}^T \sqrt{\sum_{k=1}^K \beta_{t,k}^2} \quad (6)$$

is penalized by a penalization parameter  $\lambda_1$  in order to shrink parameter vectors groupwise towards zero. The last component is the matrix notation of the Adaptive Group Fused LASSO (Zou, 2006) which penalizes parameter vector changes over time. The adaptive weighting vectors  $\dot{w}$  help to improve the asymptotic estimation properties of the AGFL w.r.t. variable selection <sup>2</sup>. To do so, the data-driven weights  $\dot{w}$  account for grouped parameter vector changes reversely to

<sup>2</sup>See oracle properties of adaptive LASSO (AL) in Appendix A.2

its  $\ell_{2,1}$ -norm of parameter vector changes generated by OLS pre-estimates  $\dot{\beta}_t$ :

$$\dot{w} = \left( \begin{array}{c} \|\dot{\beta}_2 - \dot{\beta}_1\|^{-\kappa} \\ \vdots \\ \|\dot{\beta}_2 - \dot{\beta}_1\|^{-\kappa} \\ \|\dot{\beta}_3 - \dot{\beta}_2\|^{-\kappa} \\ \vdots \\ \|\dot{\beta}_T - \dot{\beta}_{T-1}\|^{-\kappa} \end{array} \right)' \quad \left. \vphantom{\begin{array}{c} \|\dot{\beta}_2 - \dot{\beta}_1\|^{-\kappa} \\ \vdots \\ \|\dot{\beta}_2 - \dot{\beta}_1\|^{-\kappa} \\ \|\dot{\beta}_3 - \dot{\beta}_2\|^{-\kappa} \\ \vdots \\ \|\dot{\beta}_T - \dot{\beta}_{T-1}\|^{-\kappa} \end{array}} \right\} K \quad (7)$$

$_{[1 \times (T-1)K]}$

Intuitively, the size-inverse weighting can be understood as a compensating correction for the dominance of larger parameter vector changes in terms of the  $GTV$ . Due to the balancing character of the adaptive weights, the optimization algorithm is prevented from focusing predominantly on periods with relatively high  $GTV$ . Hereby, the weighting parameter  $\kappa$  determines the preferences for balancing the dominance of relatively low parameter vector changes. Altogether, the objective of optimization problem (5) is to segment time intervals with constant dynamics by fusing parameter vector over time in view of the trade-off between model fit and parameter sparsity and parameter stability. Equation (5) is then optimized for each single county in order to detect change points on the micro-level.

As proposed by [Alaíz et al. \(2013\)](#), the optimization problem (5) is solved via the FISTA algorithm, ([Beck and Teboulle, 2009](#)) which is an extension of the basic gradient descent technique to the case of optimization problems including a regularization term.<sup>3</sup> Importantly, the number of detected change points  $J - 1$  strongly depends on the choice of the penalization parameters  $\lambda_1$  and  $\lambda_2$  in (5). Hereby,  $\lambda_1$  penalizing the absolute parameter size is set to 0, as shrinking absolute parameter values seems to affect the ability of the AGFL algorithm to shrink parameter change. Moreover,  $\lambda_2$  is tuned with respect to the target to report 10% of the periods as change points. As explained in the next section, the AGFL model will serve the purpose to provide parameter estimates  $\hat{\beta}_t$  and one-step-ahead predictions  $\hat{y}_{t+1}$  based on the most recent dynamics of the model. Given this purpose, it is more problematic in terms of the bias to underestimate the number of change points than to overestimate it as basing the one-step ahead prediction upon obsolete data leads to biased residuals. According to the bias-variance trade-off, having

---

<sup>3</sup>The applied algorithm is explained in [Appendix A.1](#) in more detail.

data based on a shorter segment increases the variance in the estimates, but should decrease the one-step-ahead prediction error  $e_{t+1}$ .

Eventually, the outcome of the model is the estimated parameter vector  $\hat{\beta}$  from (5) that exhibits constant parameter vector estimates  $\hat{\beta}_{i,t}$  for the periods  $t \in \{T_j, \dots, T_{j+1} - 1\}, j \in \{1, \dots, J\}$ , which can be rewritten as  $\hat{\alpha}_{i,j} = \hat{\beta}_{i,T_j} = \hat{\beta}_{i,T_{j+1}} = \dots = \hat{\beta}_{i,T_{j+1}-1}$ . Hence, the time-variant regression model is smoothed by the total variation penalization generating  $J - 1$  change points marking  $J$  homogeneous segments with constant parameter estimates  $\hat{\alpha}_{i,j}, j = 1, \dots, J$ .

Furthermore, this historic change point information is denoted by  $cp.mic_{i,t}^{AGFL} = 1$  resp.  $-1$ , where the sign of the change point indicator corresponds to the sign of the one-step-ahead prediction error  $e_{i,t} = y_{i,t} - \hat{y}_{i,t}$  associated with a significant parameter change  $\hat{\beta}_{i,t} - \hat{\beta}_{i,t-1}$ . Eventually, the residuals  $e_{i,t}$  provide fundamental information about the direction and the strength of a change in the underlying model.

The parameter vector of the latest identified segment  $\hat{\alpha}_{i,J}$  is used to predict the dependent variable for the next period  $\hat{y}_{i,t+1}$ . In the absence of significant change points, the estimated parameter vector estimate  $\hat{\alpha}_{i,j}$  will be similar for the subsequent periods and hence, the residuals  $e_{i,t}$  based on the AGFL estimates reflecting the latest dynamics are expected to randomly fluctuate around 0. If the dynamics in the pandemic development represented by (21) change, the one-step-ahead predictions  $\hat{y}_{i,t+1}$  will no longer be close to the observed  $y_{i,t+1}$ . Instead, in case of an upswing in the pandemic, the predictions  $\hat{y}_{i,t+1}$  reflecting the previous development of the pandemic will be lower than the observed values  $y_{i,t+1}$  and the residuals are expected to turn positive during the transition period. Analogously, the one-step-ahead residuals are expected to turn significantly negative in case of a substantial downswing in the spread and strength of a pandemic. Hence,  $cp.mic_{i,t}^{AGFL} = 1$ , if the residual  $e_{i,t}$  is positive and  $cp.mic_{i,t}^{AGFL} = -1$  in the opposite case. In the comparative analysis, these retrospectively identified CPs will serve as reference CPs for the online CPD models.

### 2.3 Online CPD

The previously introduced AGFL (5) refers to the task of offline CPD. Solving the optimization problem with tuned hyper-parameters directly generates change points that delineate periods with constant parameter vector estimates. In order to adapt the AGFL estimator to an online

CPD method, the model must be extended to process real-time data with the objective to assess whether the incoming data indicate changes in the dynamics of the pandemic. The first approach to this is the aggregated residual alarm (ARA) model.

Let the macro-level *alarm statistic*  $M_t$  be defined by the law of motion

$$M_t = \lambda M_{t-1} + (1 - \lambda)(e_t / \hat{\sigma}_{t-1}) \quad (8)$$

where  $e_t = y_t - \hat{y}_t$  is the aggregated one-step-ahead residual and  $\hat{\sigma}_t$  is the sample residual variance. As the weighted sum of a rolling window of residuals, the alarm statistic  $M_t$  is substantially affected by the smoothness parameter  $\lambda$  that governs the dominance of historic residuals. Hence,  $\lambda$  can be interpreted as a discount factor for the historic residuals which have a decreasing impact on  $M_t$  that can be defined as the geometric series

$$M_t = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j \frac{e_{t-j}}{\hat{\sigma}_{t-1-j}}. \quad (9)$$

The alarm statistic contains information about the previous prediction errors of the AGFL and, hence, is a measure for the magnitude of the change in the latest dynamics of the pandemic. As soon as the alarm statistic exceeds a threshold value  $\kappa$ , a change point on the macro-level is reported in the following way:

$$cp.mac_t^{ARA}(\kappa) = \begin{cases} 1 & , \text{ if } M_t \geq \kappa \\ 0 & , \text{ if } -\kappa < M_t < \kappa \\ -1 & , \text{ if } M_t \leq -\kappa \end{cases} \quad (10)$$

The decision rule (10) yields a ternary-coded, aggregated CP indicator signalling whether a CP exists at time  $t$  and - in a case of a CP ( $cp.mac_t^{ARA} \neq 0$ ) - whether the reported CP is an up- ( $cp.mac_t^{ARA} = 1$ ) or downswing ( $cp.mac_t^{ARA} = -1$ ). In this ternary-choice model, the two parameters  $\lambda$  and  $\kappa$  are selected according to an objective function. The loss

$$(\lambda^*, \kappa^*) = \underset{\lambda, \kappa}{argmin} \underbrace{\alpha FN^+ + (1 - \alpha) FN^-}_{\text{missed change points}} + \underbrace{\beta FU^0 + (1 - \beta) FD^0}_{\text{false alarms}} + \underbrace{\phi FD^+ + (1 - \phi) FU^-}_{\text{contrary alarms}}, \quad (11)$$

serves as a flexible objective function allowing for a flexible weighting of falsely reported change points in the sense of cost-sensitive learning. Eventually, in contrast to the classical binary classification problem, the considered key variable  $cp.mac_t^{ARA}$  is ternary and hence, the confusion matrix and the loss structure for false predictions have to be tailored to a more sophisticated classification set-up with three classes. The notational convention is given by Table 1.

<b>Predicted/Actual</b>	$cp.mac_t^{AGFL} = 1$	$cp.mac_t^{AGFL} = 0$	$cp.mac_t^{AGFL} = -1$
$cp.mac_t^{ARA} = 1$	$TU$	$FU^0$	$FU^-$
$cp.mac_t^{ARA} = 0$	$FN^+$	$TN$	$FN^-$
$cp.mac_t^{ARA} = -1$	$FD^+$	$FD^0$	$TD$

Table 1: Count variables for (Predicted/Actual)-pairs in a 3x3 confusion matrix. The first letter reveals, whether the prediction is true or false. The second letter represents the predicted category and in case of a false prediction, the superscript depicts the actual category.

It is important to mention that the indicator variables in the confusion matrix depict count variables of the corresponding (Predicted/Actual)-pairs during the complete observation period. Moreover, the presented confusion matrix can be calculated w.r.t. macro- or micro-level reference CPs. In case of the micro-level, the count variables in the confusion matrix represent how well the macro-level ternary variable  $cp.mac_t^{ARA}$  fits the micro-level CPs ( $cp.mic_{i,t}^{AGFL}$ ). In case of the macro-level, the indicator variables of the confusion matrix are based on the reported CP variable  $cp.mac_t^{ARA}$  in comparison to the reference macro-level CPs from the AGFL ( $cp.mac_t^{AGFL}$ ).

The tuning parameters  $(\alpha, \beta, \phi)$  have to satisfy some constraints following a *reasonable* loss pattern in the sense of cost-sensitive learning. Clearly, costs of failing to report an upswing are higher compared to a downswing ( $\alpha > 0.5$ ). Additionally, it is worse to report a false downswing compared to a false upswing ( $\beta < 0.5$ ), as a false downswing could lead to premature relaxation of lockdown measures. Moreover, it is more severe to miss an upswing than to falsely report an upswing ( $\alpha > \beta$ ). Furthermore, contrary alarms ( $FD_t^+, FU_t^-$ ) are more malicious than the fail to report a change point ( $FN_t^+, FN_t^-$ ) and than false alarms in situations without real change point ( $FD_t^0, FU_t^0$ ), i.e.  $\phi > (1 - \beta)$  and  $\phi > \alpha$ . Lastly, contrary prediction in case of a true upswing ( $FD_t^+$ ) is worse than in case of a true downswing ( $FU_t^-$ ), i.e.  $\phi > 0.5$ .

The parameter ratios can be interpreted reasonably.  $\alpha/\beta$  expresses the aversion of missing an upswing compared to reporting an upswing if there is no real change point. Moreover,  $\alpha/(1 - \alpha)$  represents the aversion of the model against missing upswings relative to missing downswings.

As a starting point, given the parameter restrictions from above, I set  $(\alpha, \beta, \phi) = (2/3, 1/3, 3/4)$ . The parameters might be set reasonably by taking the real-world costs of the different types of false reports into account. Hence, based on this so-called cost-sensitive learning approach, the parameters  $(\kappa, \lambda)$  are tuned w.r.t. the objective to minimize the real-world costs of the outcome of the CPD model. Having specified the objective function (11), the hyper-parameters  $(\kappa, \lambda)$  are then tuned via a dynamic, two-dimensional coarse-to-fine grid search as explained in 2.7.

## 2.4 Model Extension: Homogeneity Measure for Micro-Level Change Points

In response to the pitfalls of the alarm statistic  $M_t$ , the goal is to derive a supplementary alarm statistic to (8) which indicates the CP homogeneity across counties. In order to derive such a measure, micro-level change points have to be detected by calculating (8) on micro-level using micro-level residuals. The identified change points  $cp.mic_{i,t}^{ARA} \in \{-1, 0, 1\}$  are used to calculate CP proportions indicating the spread of the different CP states:  $P_t^+ = \sum_{i=1}^N \omega_i \cdot \mathbb{1}(cp.mic_{i,t}^{ARA} = 1)$ ,  $P_t^0 = \sum_{i=1}^N \omega_i \cdot \mathbb{1}(cp.mic_{i,t}^{ARA} = 0)$ ,  $P_t^- = \sum_{i=1}^N \omega_i \cdot \mathbb{1}(cp.mic_{i,t}^{ARA} = -1)$ . As the proposed supplementary measures follow the DTA principle, the ARA is extended to a mixture approach combining the information value incorporated in ATD and DTA measures. A first DTA measure is the total number of detected change points  $cp.share_t$  and the difference of micro-level up- and downswings  $cp.excess_t^+$ .

$$cp.share_t = P_t^+ + P_t^- . \quad (12)$$

$$cp.excess_t^+ = P_t^+ - P_t^- . \quad (13)$$

The difference measure  $cp.excess_t^+$  still can not distinguish scenarios of 'no change points' and 'equally many up- and downswings'. To account for the heterogeneity in the micro-level data, the excess share of upswings over downswings  $cp.excess_t^+$  is divided by the total share of identified change points  $cp.share_t$ :

$$cp.homogeneity_t = \frac{cp.excess_t^+}{cp.share_t}, cp.share_t > 0. \quad (14)$$

The measure  $cp.homogeneity_t$  is easily interpretable as it is bounded between  $[-1, 1]$ , where, given  $cp.share_t > 0$ ,  $cp.homogeneity_t = 0$  represents perfect heterogeneity and  $|cp.homogeneity_t| = 1$  perfect homogeneity across all single counties. However,  $cp.homogeneity_t$  will still exhibit an am-

ambiguous statement if  $P_t^+ = P_t^-$ . However, as soon as  $P_t^+$  and  $P_t^-$  differ slightly, *cp.homogeneity<sub>t</sub>* will be able to differentiate between situations with many and only few total change points.

## 2.5 Model Evaluation Measures

Based on the 3x3-confusion matrix (Table 1), categorical evaluation criteria analogous to the binary classification case can be derived (acc. to Sokolova and Lapalme (2009)):

$$Prec^+(Upswings) = \frac{TU}{TU + FU^0 + FU^-} \quad (15)$$

$$TPR^+(Upswings) = \frac{TU}{TU + FN^+ + FD^+} \quad (16)$$

$$FPR^+(Upswings) = \frac{FU^0 + FU^-}{TN + TD + FD^0 + FN^- + FU^0 + FU^-} \quad (17)$$

$$Acc(Total) = \frac{TU + TN + TD}{N} \quad (18)$$

Auxiliary 2x2-confusion matrices are helpful to determine the class specific evaluation criteria  $Prec^+$ ,  $TPR^+$  and  $FPR^+$ . These 2x2-confusion matrices are obtained by merging the two negative classes to the negative class respectively. The resulting confusion matrices are stated in Appendix A.4. It is important to mention that the precision and the recall are only class-specific measures while the accuracy is a measure for the prediction performance over all categories. Hence, the precision and recall can analogously be calculated of neutrals or downswings. Defining the recall and the precision makes the definition of the specificity redundant as it is equivalent to the recall w.r.t. downswings.

Moreover, the classification performance is evaluated by the ROC- and AUC-concept. The ROC-curve depicts the pairs of recall and false-positive-rate achievable by a classifier. In case of three categories, ROC-curves can be constructed for each category (positive, neutrals, negatives). The three corresponding AUC-values yield a dense performance measure that can be used to compare CPD models. In this application, I will primarily analyse the ROC curve and its corresponding AUC of the class of 'neutrals' against 'non-neutrals' (i.e. CPs) in order to analyse the models ability to discriminate between CPs and neutral periods in general. The ROC representation of the class 'neutral' contrasts the neutral periods (no) with CP periods (up- and downswings), i.e. the FPR is the share of missed CPs relative to the number of CPs and the TPR can be interpreted as the share of correctly identified 'neutrals' which is 1 minus the share

of false alarms relative to the number of all 'neutrals'.

One important caveat regarding the presented confusion matrix and the derived evaluation criteria is that all the measures are strict in the sense that these measures do not account for the detection lag of a CP, i.e. a missed CP is considered in the confusion matrix independent of the length of the detection lag. Another problem associated with the proposed micro-level measures is that the evaluation measures (15)-(18) strongly depend on the heterogeneity of the micro-level CPs across Germany. If the CP dynamics are highly heterogeneous, a single country-level CP indicator (10) will naturally yield a higher share of misclassifications which impedes the comparison of micro- and macro-level measures and the comparison of micro-level measures for different samples.

## 2.6 Non-trivial benchmark online CPD models

### 2.6.1 Change Point $\tau$ -Test

The change point  $\tau$ -test (CPTT) (see Lütkepohl (2005), Chapter 4.6.2) can be used as a non-trivial benchmark online CPD model for the previously proposed ARA. Under the  $H_0$  of  $y_{T+1}$  being generated from the prior model dynamics  $\hat{\alpha}_J$ , the normality assumption of the residuals  $e_t \sim N(0, \Sigma)$  is used to pin down a well-defined distribution of the test statistic which serves as a measure for the deviation from the prior model dynamics at time  $T$ :

$$\tau = e_T^2 / \hat{\Sigma} \sim \chi^2(1), \hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (e_t - \bar{e}_t)^2 \quad (19)$$

Based on the test decision, a change point at  $T$  is reported if the test statistic  $\tau > c_q$  where  $c_q$  is the  $q$ -quantile of the  $\chi^2(1)$ -distribution. As stated for the ARA in the previous section, the direction of the change can be deduced from the sign of the AGFL one-step-ahead residuals  $e_T$ . To check the validity of the underlying normality assumption of the test, I employ the Jarque-Bera-Test (Jarque and Bera, 1980) to see how well the third and fourth moment of the empirical residual distribution matches the corresponding theoretical counterparts of the normal distribution.

After checking the normality assumption, the CPTT can be performed on the micro- or macro-level based on the corresponding estimated residual variance  $\hat{\Sigma}$  in order to obtain micro- and macro-level CPs  $cp.mic_t^{CPTT}$  and  $cp.mac_t^{CPTT}$ . Moreover, the DTA measures stated in (12)-(14) are calculated based on the micro-level CPs detected  $cp.mic_t^{CPTT}$ . Analogously to the

alarm statistics of the ARA, the homogeneity measure  $cp.homogeneity_t^{CPTT}$  is calculated for the whole country.

### 2.6.2 Macro-only Residual Alarm (MORA) Model

In order to demonstrate the relevance of micro-level data for the task of CPD, I set up an online CPD model having the same structure as the ARA which is agnostic to micro-level data. Hence, the MORA model just relies on AGFL estimates (5) of the macro-level and the alarm statistic  $M_t$  (8) directly calculated based on macro-level residuals  $e_t$  instead of the aggregated micro-level residuals  $e_{i,t}$ . Based on the tuned ternary-choice decision model (10) the macro-level CPs  $cp.mac_t^{MORA}$  are reported. Obviously, there exist no DTA measures if the micro-level data are neglected.

## 2.7 The Model Procedure: Connecting the Model Components

The complete model procedure is depicted in the pseudo-code `detect.CP` in Appendix A.3. CPs are monitored in a real-time manner at a daily frequency. The rolling window has a length of  $h = 52$  days and is separated into a tuning period of the first  $h_1 = 40$  days and a monitoring period of the subsequent  $h_2 = 12$  days:

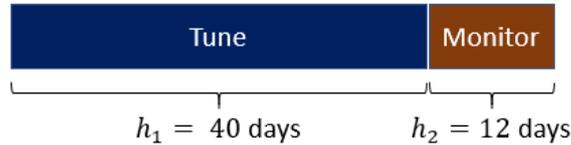


Figure 1: Separation of the rolling observation period into a tuning and monitoring sample.

Initially, `estimate.AGFL` (in Appendix A.3) is run for the observation period in order to obtain one-step-ahead residuals  $e_{i,t}$  resp.  $e_t$  and reference CPs on the macro-level. The remaining pseudo-code can be intuitively partitioned into three sections:

#### 1. Tuning:

Based on the tuning period of 40 days, a coarse-to-fine grid search algorithm on the ATD model (`run.ATD` in Appendix A.3) is run in order to find the optimal parameters  $(\lambda^*, \kappa^*)$ . The objective function (11) is used to evaluate the model.

## 2. Monitoring:

Subsequently, the alarm statistic (8) is calculated for the monitoring period. Given the optimal parameters  $(\lambda^*, \kappa^*)$ , a CP is reported according to decision rule (10). The dynamic tuning of the parameters ensures that the decision model based on the alarm statistic (8) and the decision rule (10) may continuously adapt to environmental changes over time.

## 3. Model evaluation:

According to Section 4, the ARA is tested based on a monitoring period of 60 days in order to obtain more trustable performance measures. For this purpose, the sample is extended to 100 days in which the tuning and monitoring procedure is run 49 times (49 possible 12-day-periods within 60 days), i.e. `detect.CP` (in Appendix A.3) is run 49 times for  $t = T - 48, \dots, T$ . Retrospectively, the reported CPs by ARA (`run.ATD`) can be compared to the reference CPs detected by AGFL (`estimate.AGFL`) and evaluated by the measures in 2.5.

As an extension, `get.DTA.measures` yields a dense measure (cp. (14)) for the degree of homogeneity in the pandemic dynamics across the German counties.

### 3 Data

In this work, the application is to detect CPs in the epidemic development of the current COVID-19 infection cases in Germany. In order to capture the current state of the pandemic, I use the time series of new COVID-19 infections. In more detail, I use panel data provided by the *Robert-Koch-Institut*<sup>4</sup> containing information about COVID-19 infections at a daily frequency for the most granular level available, i.e. for each of the 401 German counties. Concretely, the dependent variable representing the current infection activity within county  $i$  is a moving average measure relative to the population size, which is also known as the *seven-days incidence*:

$$y_{i,t} = \frac{1}{7} \sum_{m=0}^6 \frac{NewInfections_{i,t-m}}{Population_i} \cdot 1000000. \quad (20)$$

Choosing the *seven-days incidence* is reasonable for two good reasons. First, correcting for the population size enables direct inter-county comparisons (see e.g. Figure 5(g)) and secondly, even more important, the seven-days average gets rid of weekly seasonalities in the testing patterns and registration routines that may drive daily infection numbers substantially. Hence, the dependent variable  $y_{i,t}$  is free of seasonality. Some figures of the infection dynamics are provided in Appendix A.5.1. The data set on the COVID-19 infections is available on county level (*Land- and Stadtkreise*). Moreover, the data set contains lockdown variables ( $MOB_{i,t-7}$  in Eq. (21)) indicating the populations' movement behaviour that might be an important driver of the infections. As a default lockdown state dummy variable is not reasonable in view of the heterogeneous and quickly changing lockdown measures, I proxy the lockdown states by movement data of the population from the mobility report<sup>5</sup> provided by *Google*. Eventually, it seems obvious that the mobility of the population is a good control variable for the infection dynamics in the close future. In order to capture the natural delay in the real effect and to avoid endogeneity problems I use the 7-day lag of the mobility information on federal state level (the highest detail level available). The mobility report contains a measure of relative mobility of the population against a reference day<sup>6</sup> w.r.t. different location categories or resorts of the

---

<sup>4</sup>[www.npgeo-corona-npgeo-de.hub.arcgis.com](http://www.npgeo-corona-npgeo-de.hub.arcgis.com), state: 27.10.2020

<sup>5</sup>Google LLC "Google COVID-19 Community Mobility Reports". <https://www.google.com/covid19/mobility/> Accessed: 2020/11/19.

<sup>6</sup>The reference day is defined to be respective weekday mean of travelled distance in the period between 3<sup>rd</sup> of January and 6<sup>th</sup> of February 2020.

daily life. For this application, I decided to use the mobility data for the resort 'Retail and Recreation' incorporating the activity of the population in shopping zones, museums, cinemas and restaurants as a proxy for the lockdown state as this resort might be affected most by a lockdown. Other resorts such as 'Workplaces' or 'Supermarket and Pharmacy' might react less flexible on a lockdown and, furthermore, might reflect the population's acceptance of the lockdown policies worse than the resort 'Retail and Recreation'. Apparently, people might be able to change their leisure behaviour more self-determined as their working behaviour. Some visualizations on the regressor of mobility are provided in Appendix [A.5.3](#).

Finally, the baseline model underlying the CPD model is

$$y_{i,t} = \begin{pmatrix} 1 & y_{i,t-1} & MOB_{i,t-7} & t_{i,t} \end{pmatrix} \begin{pmatrix} \beta_{i,1} \\ \beta_{i,2} \\ \beta_{i,3} \\ \beta_{i,4} \end{pmatrix} + \varepsilon_i. \quad (21)$$

where  $y_{i,t-1}$  is the lagged dependent variable,  $MOB_{i,t-7}$  indicates the mobility of the population. Moreover, a deterministic trend  $t_{i,t}$  and an intercept is included in the model equation.

## 4 Results

In the following, the real-time CPD model ARA elaborated in Section 2.3 is evaluated with respect to their ability to date change points and. Moreover, the ARA is compared with the naive CPTT presented in 2.6.1 and the MORA, which is agnostic to micro-level data. Due to the lack of commonly accepted CPs in the incidence of COVID-19 infections in Germany, the CPs detected by the offline CPD model AGFL are employed as reference CPs. As evaluation criteria, I take detection lags and the quality measures (15), (18) and the AUC measure into account.

### 4.1 The reference change points

The AGFL (5) detects dates of positive (upswings) and negative (downswings) change points. During the period from 8<sup>th</sup> of October 2020 until 15<sup>th</sup> of January 2021, 596 individual upswings and 960 downswings have been reported (see Figure 8). In the pre-training period from 8<sup>th</sup> of October 2020 to 15<sup>th</sup> of November, 142 up- and 214 downswings and in the considered sample period from 16<sup>th</sup> of November 2020 to 15<sup>th</sup> of January 2021, 454 up- and 746 downswings have been reported. Plotting the time series of daily reported change points reveals valuable information about the historic dynamics of the COVID-19 infections. In the Figures 9 and 10, I consider the 5%-quantile resp. the 10%-quantile of the number of individual up- and downswings as candidates for national change points. The reference CPs based on the different quantiles are noted in Tables 6 and 7. The two quantiles reflect different significance levels for change points. In the observation period ('2020-11-16' - '2021-01-15'), based on the 5%-quantile-criterion, the AGFL reports upswings at '2020-12-07' (together with '2020-12-08' and '2020-12-09') and on '2021-01-06' and downswings at '2020-12-22' (together with '2020-12-23' and '2020-12-24') in the COVID-19 infections (see Fig. 9). Moreover, the 10%-quantile CPs also include a downswing at '2021-01-11'.

### 4.2 Online CPD model

In contrast to the retrospective change point detection method used as a reference, the proposed online CPD models have to face the more challenging task of identifying change points in real-time. The latter approach is more demanding as, at each time, the change point classification has to be performed without having the future observations at hand. In Figure 12, up- and

downswings detected by ARA are marked in the graph of the incidence of COVID-19 infections  $y_t$ . The optimized ARA  $((\lambda^*, \kappa^*) = (0.9, 1.5))$  reports the first upswing ('2020-12-07'-'2020-12-09') with a delay of three days on '2020-12-10'-'2020-12-13'. The second upswing (on '2021-01-06') is reported two days too late. The AGFL downswing on '2020-12-22'-'2020-12-24' is reported with a delay of three days. The downswing '2021-01-13' following the reference CP definition based on the 10%-quantile is not reported by the end of the observed period ('2021-01-15'). The ARA reports all relevant CPs (acc. to 5%-quantile) with a delay of one day at maximum and does not alert falsely. Generally, a potential reporting lag might result from the fact that the moving sum of residuals of the previous 10 days  $M_t$  still contains aggregated residuals of the *old* regime before the change point happens. Especially if the residuals in the *old* regime point in a contrary direction, i.e. if the aggregated residuals  $e_t$  in (8) are negative before an upswing or positive before a downswing, the alarm statistic will exceed the threshold parameter  $\kappa$  too late. As discussed in Section 2.3, the alarm statistic  $M_t$  (8) might fail to distinguish situations with dispersed pandemic development from situations without any change point, such that the ATD alarm statistic is only meaningful, if the micro-level units develop fairly synchronously. In this context, the DTA measure  $cp.homogeneity_t$  (14) might help to distinguish both scenarios.

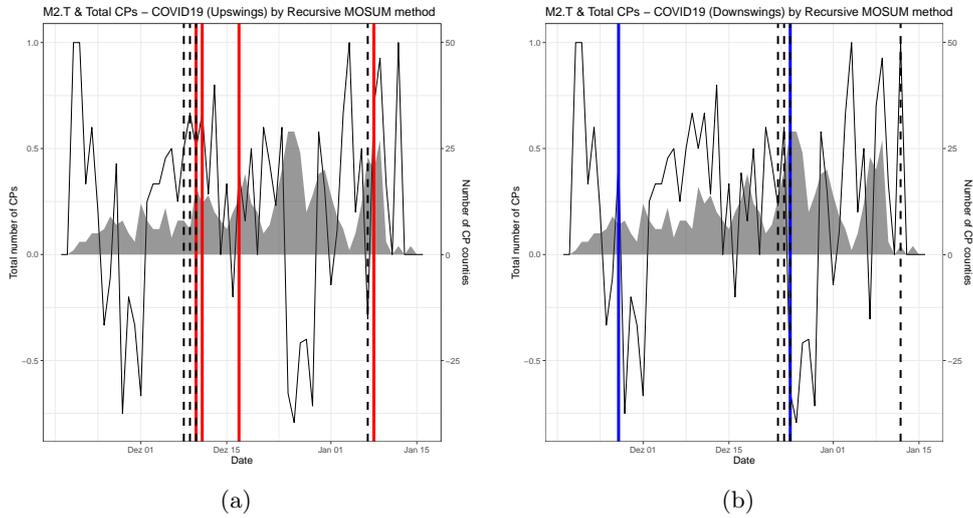


Figure 2: Reported (a) upswings and (b) downswings based on model ARA using a monitoring window length of 10 in the plot of the homogeneity measure  $cp.homogeneity_t$  (solid line) and the count of reported CPs  $cp.share_t$  (shaded area). The reference change points are marked as dashed vertical lines respectively, whereas the coloured vertical lines depict the up- and downswings detected by ARA.

Figure 2 visualizes the homogeneity measure  $cp.homogeneity_t$  and the share of reported

change points  $cp.share_t$  together with the reported change points. Both DTA measures reveal details from the micro-level which help to interpret the binary CP information correctly. The homogeneity measure helps to detect periods of heterogeneously changing dynamics across all counties, when the ATD fails to detect a national CP. The measure  $cp.share_t$  indicates the spread and the CP prevalence in Germany. The analysis of the supplementary DTA measures reveals that the secondly reported downswing in Figure 2 results from much more micro-level change points (29 vs. 8). At the same time, the homogeneity measure exhibits similar values for both days (-0.793 vs. -0.750) indicating a very high degree of homogeneity compared to the remaining observation period.

On the one hand, the two proposed measures provide more detailed information about the reported change points w.r.t. the homogeneity ( $cp.homogeneity_t$ ) and the spread ( $cp.share_t$ ) of the change points on the micro-level.

On the other hand, scenarios of many contrary change points and of no change points on the micro-level can be easily distinguished by monitoring the two presented measures even if no change point is reported. Hence, this analysis serves as a powerful supplement to monitoring the alarm statistic  $M_t$ . When analysing the infection dynamics at and two weeks after Christmas, it is clear that the incidence is strongly driven by the reduced COVID-19 tests and an irregular reporting routine of the local health authorities (*Gesundheitsämter*). Clearly, based on the reported infections, the model does a good job in detecting a downswing on '2020-12-24' and an upswing on '2021-01-08'. However, the identified upswing is likely to be artificially generated by late registrations from the period between Christmas and Epiphany <sup>7</sup>.

### 4.3 Non-trivial benchmark CPD models

#### 4.3.1 CPTT

As introduced in Section 2.6.1, the CPTT is a real time change point test based on one-step-ahead forecast residuals. As the CPTT assumes normality of the residuals under the  $H_0$ , the Jarque-Bera-test results are reported in Table 2. The Jarque-Bera-test clearly rejects the normality of the empirical distribution of the residuals w.r.t. the third and the fourth moments. A visual check can be performed to capture the most salient aspects of the non-normality in the

---

<sup>7</sup>'2021-01-06': Public holiday in most of the German federal states

Type	Statistic	df	p-value
Normality	68.993	2	0.0000
Skewness	0.665	-	0.0066
Kurtosis	6.846	-	0.0000

Table 2: Jarque-Bera-test results suggest rejecting  $H_0$  of normal residuals. Clearly, the non-normality applies to both, the third and fourth, moments of the empirical distribution of the residuals.

residuals (see Appendix A.7.2). According to the Jarque-Bera-test results (2), the skewness of the empirical distribution of the residuals is not symmetric due to some outliers and a slightly left-shifted peak. Moreover, the kurtosis seems to fit better to the normal distribution, but is still rejected by the Jarque-Bera-test. The comparably broad peak and the comparably strong tails are non-normal. Hence, the normality assumption is too restrictive implying that the CPTT is a sub-optimal choice for detecting CPs in the given data set. Though, this naive CP test yields a useful non-trivial benchmark for the ARA.

Figure 13 visualizes the CPs reported by the CPTT. On the one hand, the CPTT lately reports the upswings from '2020-12-07'-'2020-12-09' with a delay of 9 days. Moreover, a false alert of an upswing is given at '2020-12-23' when a downswing occurred (acc. to AGFL reference). The upswing at '2021-01-06' is reported by the CPTT two days too late. On the other hand, the CPTT reports the downswings between '2020-12-22' and '2020-12-24' at two days too late and is the only model detecting the *weak* (only detected by 10% quantile AGFL) downswing at 13<sup>th</sup> of January 2021 with a delay of 4 days. Moreover, the change point test falsely reports a downswing at '2020-11-30'.

### 4.3.2 Macro-only model

Theoretically, making use of micro-level information might bear valuable insights. Hence, this comparing the ARA with its macro-only counterpart MORA quantifies the loss induced by neglecting the micro-level data.

As visualized in Figure 15, the MORA reports an upswing at '2020-12-06' already one day in advance. Moreover, an upswing is identified at '2020-12-16', i.e. the first reference upswing (5%-quantile) is hit inaccurately. The reference upswing at '2021-01-06' is reported only one day too late. Moreover, some false alerts are reported on '2020-11-22' and '2020-12-31'.

The downswing at '2020-12-22' is reported two days too late. Moreover, three false downswings

are reported at '2020-11-24' and '2020-11-30' and '2021-01-06'. In total, the MORA yields a diffuse picture of the reference CPs characterized by many false alarms compared to the ARA including micro-level information.

#### 4.4 Comparison

Figure 16 (Appendix A.7.2) visualizes the CPs reported by the ATD ( $(\lambda^*, \kappa^*) = (0.9, 1.5)$ ) together with the benchmark models, the CPTT (critical value: 5% quantile) and the MORA ( $(\lambda^*, \kappa^*) = (0.7, 1)$ ), alongside the reference CPs detected by the AGFL (5% quantile).

The first upswing at '2020-12-07' is firstly reported by the MORA one day in advance. The ARA reports this upswing with a delay of three days and the CPTT with a delay of nine days. The upswing at '2021-01-06' is reported by the ARA and CPTT simultaneously two days too late. The MORA just misses the AGFL upswing by one day. Moreover, the CPTT gives one and the MORA two false alerts.

The downswing in the infection dynamics at '2020-12-22'-'2020-12-24' is detected by the MORA and the CPTT one day too late ('2020-12-23'). The ARA even reports this downswing even one day later. The CPTT detects the downswing at '2021-01-12' with a comparably small delay of 3 days. Moreover, the change point test does not falsely report a downswing at '2020-11-30' like ARA. The MORA usually seems to report CPs one day earlier than the ARA. Moreover, the MORA (optimized w.r.t. objective function (11)) seems to react more sensitive and less discriminative to extreme residuals as it reports some CPs which are not reported by the ARA. As Figure 16 illustrates is the MORA the model which reports the reference CPs first with a reporting delay of 2 days at maximum. The ARA has a reporting lag of less than 3 days. The CPTT reports the reference CPs with a delay of up to nine days. However, in contrast to the MORA and the CPTT, the ARA never falsely reports a change point in the monitoring period. This visual analysis indicates that the ARA is reporting more trustable compared to MORA and CPTT. The ARA is the only model which reports all up- and downswing with a delay of 3 days at maximum and, moreover, never alerts falsely. From a qualitative point of view, the ARA clearly outperforms the other two non-trivial, tuned benchmarks CPD models having the highest sensitivity and specificity among the proposed models. It hits all reference CPs with a small delay (at maximum 3 days) and does not alert falsely.

The qualitative, visual analysis shall be supplemented by the comparison of the quantitative evaluation criteria from Section 2.5 to enable a strict comparison of the different models. Though, the evaluation criteria still have to be taken with caution as these measures just consider the classification into 'upswing', 'downswing' or 'neutral' day by day neglecting the temporal distances between different days. Hence, the classification accuracy (18) will be the same no matter whether a CPD model misses the true CP by one or several days. The evaluation measures are blind w.r.t. the aspect of the reporting lag of the CPD models. Put differently, the evaluation criteria are incomplete instrumentalization of the CPD detection performance.

Moreover, it is difficult to find a commonly accepted overall performance measure. Given the model performance criteria of reliability and timeliness of the reported CPs it is somehow arbitrary how to balance the trade-off between a larger prediction lag with a higher specificity.

The confusion matrices and the evaluation criteria for the ARA, the CPTT and the MORA are reported in Appendix A.8. The prediction shares (Table 14) give a first descriptive overview over the CP reporting of the different online CPD models compared to the AGFL. The MORA and the CPTT report upswings and downswings in 7% resp. 8% of the periods. The prediction pattern of the ARA is less balanced. The ARA predicts upswings in 12% and downswings in 8% of the periods.

The confusion matrices are evaluated by the *True-Positive-Rate* (TPR) and the *False-Positive-Rate* (FPR) which are measures indicating how sensitive (TPR) and exact (FPR) upswings, neutrals and downswings are detected by the real-time CPD models. Tables 16-18 reveal that, given the optimized models, the ARA predicts upswings best having the highest TPR and lowest FPR at the same time. According to the FPR and TPR, The CPTT predicts neutrals and downswings best.

The simple accuracy measure (18) suggests going for the CPTT. However, the accuracy is not a good performance measure due to an imbalanced class distribution (many neutrals). Classifiers which tend to predict the neutral class will naturally have a higher accuracy. Hence, a better evaluation criterion is the accuracy of a classifier for a fixed rate of neutral predictions or the AUC measures that takes prediction performance for different rates of neutral predictions (*positive rate*). Moreover, the accuracy measure evaluates each day separately neglecting reporting lags. The performance of the CPD models can be compared by the ROC curves (Figure 17). The ROC

curves together with its corresponding AUC-measures are performance measurements which evaluate the classification performance in terms of the FPR in dependence on the sensitivity (TPR). Based on the AUC, the ARA ( $AUC_{ARA} = 0.5476$ ) seems to beat the benchmark models MORA ( $AUC_{MORA} = 0.3452$ ) and the CPTT ( $AUC_{CPTT} = 0.4487$ ). The MORA and CPTT exhibit AUC scores lower than 0.5 meaning that the classifiers are worse than random guessing of the classes 'neutrals' and 'non-neutrals'. By accounting for the option to invert the model classification pattern,  $AUC^{effective} = \max(AUC, 1 - AUC)$  can be achieved. The effective  $AUC_{CPTT} = 0.5513$  and  $AUC_{MORA} = 0.6548$  are higher than  $AUC_{ARA}$ . This picture does not align with the qualitative analysis of the CP reports. The sharp categorization of the confusion matrix into false and true reports does not represent the timing of CPs adequately. Consequently, the limitations of the simple classification evaluation measures also apply to the ROC curve and the corresponding AUC measure.

The performance measures still strongly depend on the definition of the reference AGFL CPs (5%- or 10%-quantile). Appendix A.8.2 contains the visualizations and tables analogously to the graphs discussed before given the AGFL reference CPs based on the 10%-quantile. Moreover, the performance measures for the ARA with changing threshold parameter  $\kappa$  are provided in this section. Table 30 shows that the decrease in the prediction share of CPs improves the accuracy for the ARA when setting the threshold parameter  $\kappa = 2$ . The comparison confirms that the less sensitive models having a higher prediction rate of neutrals tend to have a higher accuracy. The lower accuracy of the ARA compared to MORA and CPTT is mainly caused by a more sensitive reporting scheme of change points. The difference in the accuracy between the different CPD models vanishes as the sensitivity of the ARA is decreased ( $\kappa = 2$ ) to the same rate of reported CPs. However, as explained in Section 2, the previous analysis has to be interpreted cautious due to the limited explanatory power of the performance measures.

## 4.5 Regional analysis

In the following, the results of ARA are analysed at a regional level which is helpful to generate more samples in order to obtain more trustable CPD results. Moreover, the regional analysis yields a deeper understanding about the degree of heterogeneity in the pandemic dynamics across the German regions. The 401 German counties are assigned to 4 German regions by federal

states. Hereby, I mainly consider geographical proximity of the federal states. Moreover, I try to take similarities w.r.t. Corona policies and the pandemic development into account. The exact grouping is described and visualized in Appendix [A.9.2](#).

Figure [21](#) and [22](#) contrast the up- and downswings detected by ARA ((a) and (b)) with the corresponding reference CPs detected by AGFL ((c) and (d)) in the four defined regions and Germany. One important finding is that most of the detected CPs ((a) and (b)) are shared by the four regions. The upswing in the incidence at '2020-12-09' on the country-level ((a)) is confirmed by upswings between '2020-12-08' and '2020-12-12' in all four regions. Moreover, the identified German-wide downswing at '2020-11-27' is confirmed by sub-regional downswings between '2020-11-26' and '2020-11-28'. Even more unambiguously, the national downswing at Christmas and the upswing after Epiphany, which might be mainly explained by the reduced testing and limited reporting of the health authorities, are almost perfectly matched by all regions. The German-wide downswing at '2020-12-24' is reflected by change points in the regions *East, North and West* at the same day and in the region *South* one day before ('2020-12-23'). Moreover, the reported national upswing at '2021-01-08' is confirmed by all regions at the same day.

The high degree of homogeneity across all sub-regions can be explained by the synchronously changed reporting behaviour during the period of frequent public holidays. Nevertheless, it seems apparent that there must have been a 'real' downswing in late December 2020 or early January 2021 as the incidence decreased persistently to a much lower level compared to the period in the mid of December 2020.

The CPD performance of ARA is evaluated by the criteria of timeliness and the accuracy of the reported change points w.r.t. the reference change points depicted in ((c) and (d)). Interestingly, the reference change points detected by AGFL reflect a much higher degree of regional heterogeneity in the CPs, especially in the upswings. Moreover, much more change points are reported on the regional level compared to the national level as contrary CPs in different regions could cancel each other out. This demonstrates the need of a homogeneity measure as  $cp.homogeneity_t$  which also detects the diversified CPs.

The analysis of the downswings ((b) vs. (d)) shows that the identified downswing in the end of November 2020 is confirmed by 'true' change points in the regions *West* and *North*. Moreover,

the downswing indicated by AGFL at Christmas is detected very timely in all regions. Lastly, the AGFL downswing at the end of the sample period ((d)) is not yet detected by the real-time CPD model ARA.

The upswings reported by AGFL exhibit a much more diffuse and heterogeneous picture compared to the downswings. However, the main structure of the reference CPs (in (c)) can be rediscovered in the CP pattern reported by ARA. On the one hand, the common upswings at '2020-12-09' and '2021-01-08' are identified by the ARA. Especially, the CPs on country-level match almost perfectly. On the other hand, some single regional upswings (e.g.the upswings in regions *North* and *East* between '2021-01-01' and '2021-01-08') are not detected by the ARA.

In total, the ARA model seems to detect the main change points very timely and reliably. However, the functionality of the ARA seems to be limited for minor change points that are not shared across all regions.

## 5 Conclusions

This study demonstrates the potential of the ATD-type online CPD models including micro-level data. The reference CPs are detected by the real-time online CPD model within a delay of three days at maximum without missing change points or reporting false alarms. One key aspect is the information value contained in the residuals of the AGFL model which accounts for structural changes in the assumed DGP. Using the micro-level data of the German counties enables the applicant to obtain a deeper understanding of the infection dynamics of Germany (macro-level). This additional information from the micro-level yields a better early-alert system (ARA compared to the MORA) and enables the analysis of infection waves w.r.t. the degree of the homogeneity.

However, one important issue of the change point analysis is that there is no common consensus about the definition of change points of the COVID-19 infection dynamics in Germany contrary to the defined business cycles according to NBER<sup>8</sup>. Hence, the evaluation of the model performance immediately depends on the choice of the reference CPs. Furthermore, the ARA relies on a large set of model specifications. By assuming a parametric DGP and change points based on the AGFL, the reported change points are sensitive w.r.t. the choice of the model set-up and the corresponding hyper-parameters.

Some aspects in context of the CPD tasks are left out for future research. In the field of panel data models, it can be fruitful to analyse change points in heterogeneous panel data taking cross-sectional information into account (Okui and Wang, 2020; Lumsdaine et al., 2020). The heterogeneous sub-entities are clustered into homogeneous sub-groups on which common model parameters are estimated. Correspondingly, the AGFL (5) could be extended to merge the parameters  $\beta_t$  not only over time, but also over the single counties in order to reduce the parameter dimensionality in the cross-sectional dimension. However, many panel models assume a constant group membership of the counties which is not appropriate in the given application. Moreover, even models allowing for a time-varying group structure will not yield a considerable improvement against the separate treatment of the counties as the group structure changes frequently (see A.5.2). Beyond that, clustering of the counties might be even malicious, when using the joint model parameters predicting future outcomes  $\hat{y}_{t+1}$ . First, using the same

---

<sup>8</sup><https://www.nber.org/research/business-cycle-dating>, Accessed: 2021/05/20

parameters for all members of a cluster might represent *extreme* cluster members inadequately as clusters might still incorporate some degree of heterogeneity. And secondly, if the clusters are obsolete or the membership of a sub-entity changes, bad predictions are caused by an outdated clustering which could lead to misleading results in the later analysis.

Moreover, it might make sense to represent the CP variable as a probability measure indicating the probability of a CP as the ternary CP representation (10) might be too strict. A continuous CP variable could express much more details about the dynamics of the pandemic.

In the scope of classification tasks along the temporal dimension there is a lot of potential in designing evaluation measures that evaluate the model performance w.r.t. time lags. In the given task, it is less important to know whether a CP is detected by the online CPD at this day, but how large the reporting lag is. Hence, the objective function should also depend on the reporting lag. The implementation of the penalization pattern of lagged predictions could be subject of a separate research project.

Furthermore, the hyper-parameters of the objective function (11) are included to allow for a cost-sensitive tuning of the CPD model. Knowing the real-world costs of false alarms or missed change points would enable the researcher to set the model parameters such that the expected real-world costs of the pandemic are minimized. Lastly, the functionality of the presented online CPD model could be systematically tested in different change point scenarios by a simulation study in the frame of predefined known change points of various magnitude and combinations of various change points.

## References

- ALAÍZ, C. M., A. BARBERO, AND J. R. DORRONSORO (2013): “Group fused lasso,” in *International Conference on Artificial Neural Networks*, Springer, 66–73.
- BACH, P., V. CHERNOZHUKOV, AND M. SPINDLER (2020): “Insights from optimal pandemic shielding in a multi-group SEIR framework,” *arXiv preprint arXiv:2011.01092*.
- BARANOWSKI, R., Y. CHEN, AND P. FRYZLEWICZ (2016): “Narrowest-over-threshold detection of multiple change-points and change-point-like features,” *arXiv preprint arXiv:1609.00293*.
- BASSEVILLE, M., I. V. NIKIFOROV, ET AL. (1993): *Detection of abrupt changes: theory and application*, vol. 104, prentice Hall Englewood Cliffs.
- BECK, A. AND M. TEBoulLE (2009): “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM journal on imaging sciences*, 2, 183–202.
- BLEAKLEY, K. AND J.-P. VERT (2011): “The group fused lasso for multiple change-point detection,” *arXiv preprint arXiv:1106.4199*.
- BUZUN, N. AND V. AVANESOV (2017): “Bootstrap for change point detection,” *arXiv preprint arXiv:1710.07285*.
- CÁSSARO, F. A. AND L. F. PIRES (2020): “Can we predict the occurrence of COVID-19 cases? Considerations using a simple model of growth,” *Science of the Total Environment*, 138834.
- CHEN, S., Q. CHEN, W. YANG, L. XUE, Y. LIU, J. YANG, C. WANG, AND T. BÄRNIGHAUSEN (2020): “Buying time for an effective epidemic response: The impact of a public holiday for outbreak control on COVID-19 epidemic spread,” *Engineering*, 6, 1108–1114.
- DAVE, D., A. I. FRIEDSON, K. MATSUZAWA, AND J. J. SABIA (2021): “When do shelter-in-place orders fight COVID-19 best? Policy heterogeneity across states and adoption time,” *Economic Inquiry*, 59, 29–52.
- DONG, C., J. GAO, O. LINTON, AND B. PENG (2020): “On Time Trend of COVID-19: A Panel Data Study,” *arXiv preprint arXiv:2006.11060*.
- GRILLENZONI, C. (2012): “Evaluation of recursive detection methods for turning points in financial time series,” *Australian & New Zealand Journal of Statistics*, 54, 325–342.
- HARCHAOUI, Z. AND C. LÉVY-LEDUC (2010): “Multiple change-point estimation with a total

- variation penalty,” *Journal of the American Statistical Association*, 105, 1480–1493.
- HAUG, N., L. GEYRHOFFER, A. LONDEI, E. DERVIC, A. DESVARS-LARRIVE, V. LORETO, B. PINIOR, S. THURNER, AND P. KLIMEK (2020): “Ranking the effectiveness of worldwide COVID-19 government interventions,” *Nature human behaviour*, 4, 1303–1312.
- HYNDMAN, R. (2020): “Forecasting COVID-19: What makes forecasting hard?” .
- JARQUE, C. M. AND A. K. BERA (1980): “Efficient tests for normality, homoscedasticity and serial independence of regression residuals,” *Economics letters*, 6, 255–259.
- JIANG, F., Z. ZHAO, AND X. SHAO (2020): “Time series analysis of COVID-19 infection curve: A change-point perspective,” *Journal of econometrics*.
- KAWAHARA, Y. AND M. SUGIYAMA (2012): “Sequential change-point detection based on direct density-ratio estimation,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 5, 114–127.
- LUMSDAINE, R. L., R. OKUI, AND W. WANG (2020): “Estimation of Panel Group Structure Models with Structural Breaks in Group Memberships and Coefficients,” *Available at SSRN 3617416*.
- LÜTKEPOHL, H. (2005): *New introduction to multiple time series analysis*, Springer Science & Business Media.
- MEINSHAUSEN, N. AND P. BÜHLMANN (2006): “Variable selection and high-dimensional graphs with the lasso,” *Ann Stat*, 34, 1436–1462.
- OKUI, R. AND W. WANG (2020): “Heterogeneous structural breaks in panel data models,” *Journal of Econometrics*.
- QIAN, J. AND L. SU (2016): “Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso,” *Journal of Econometrics*, 191, 86–109.
- SHAO, X. (2010): “A self-normalized approach to confidence interval construction in time series,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 343–366.
- SOKOLOVA, M. AND G. LAPALME (2009): “A systematic analysis of performance measures for classification tasks,” *Information processing & management*, 45, 427–437.
- STOCK, J. H. AND M. W. WATSON (2010): “Indicators for dating business cycles: Cross-history selection and comparisons,” *American Economic Review*, 100, 16–19.

- TIBSHIRANI, R., M. SAUNDERS, S. ROSSET, J. ZHU, AND K. KNIGHT (2005): “Sparsity and smoothness via the fused lasso,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 91–108.
- TRUONG, C., L. OUDRE, AND N. VAYATIS (2020): “Selective review of offline change point detection methods,” *Signal Processing*, 167, 107299.
- ZEILEIS, A. (2003): “Testing for Structural Change: Theory, Implementation and Applications,” Ph.D. thesis, Verlag nicht ermittelbar.
- ZOU, H. (2006): “The adaptive lasso and its oracle properties,” *Journal of the American statistical association*, 101, 1418–1429.

## A Appendix

### A.1 FISTA Algorithm

FISTA solves optimization problems which might have the general form

$$\min\{F(\beta) \equiv f(\beta) + g(\beta) : \beta \in \mathbb{R}^n\} \quad (22)$$

where  $f(\beta)$  represents any smooth convex function and  $g(\beta)$  any continuous convex, but potentially non-smooth function. In the given case,  $f(\beta)$  corresponds to the smooth least-squares term and  $g(\beta)$  to the non-smooth Lasso-type  $\ell_{2,1}$ -penalization terms in the AGFL optimization problem (5). The authors show that  $F(\beta)$  can be approximated at point  $\gamma$  by a quadratic function  $Q_L$  which exhibits improved convergence properties to the solution  $\beta^*$  of (22):

$$Q_L(\beta, \gamma) := f(\gamma) + (\beta - \gamma)^T \nabla f(\gamma) + \frac{L}{2} \|\beta - \gamma\|^2 + g(\beta) \geq F(\beta)^9 \quad (23)$$

with step size parameter  $L = L(f)$ .<sup>10</sup> Ignoring the constant terms w.r.t.  $\beta$  in (23) yields a unique minimum

$$p_L(\gamma) = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{L}{2} \left\| \beta - \left( \gamma - \frac{1}{L} \nabla f(\gamma) \right) \right\|^2 + g(\beta) \right\} \quad (24)$$

and the optimal  $\beta$  is recalculated

$$\beta_k = p_L(\gamma_k). \quad (25)$$

Specifically, the update steps for  $\gamma_k$  are further determined by

$$\begin{aligned} \gamma_{k+1} &= \beta_k + \left( \frac{t_k - 1}{t_{k+1}} \right) (\beta_k - \beta_{k-1}), \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2} \end{aligned} \quad (26)$$

where  $\gamma_1 = \beta_0 \in \mathbb{R}^n, t_1 = 1$ . Most importantly in this iteration scheme is that  $\gamma_k$  is computed based on the minimum of the two last iterations steps  $\beta_k$  and  $\beta_{k-1}$  (instead of traditionally only

---

<sup>9</sup> $\Leftrightarrow \|\nabla f(\beta) - \nabla f(\gamma)\| \leq L(f)\|\beta - \gamma\|$ , with  $L(f)$  being the smallest Lipschitz constant of the gradient  $\nabla f(\gamma)$   
<sup>10</sup> $L(f)$  is the smallest Lipschitz constant of the gradient  $\nabla f(\gamma)$

$\beta_k$ ), which improves the convergence properties of the algorithm substantially. As shown in Theorem (4.4) in (Beck and Teboulle, 2009), the proposed optimization algorithm will yield an  $\varepsilon$ -optimal solution at a fairly low computational complexity  $\mathcal{O}(1/k^2)$ .

## A.2 Oracle properties of the Adaptive LASSO (AL) by (Zou, 2006)

Meinshausen and Bühlmann (2006) show that variable selection of the LASSO is consistent only under some restrictive conditions. The adaptive LASSO with adaptive, data-driven weights is designed by Zou (2006) in order to establish consistent variable selection properties even under milder assumptions. In my application, consistent variable selection refers to the consistent detection of change points in the model.

Under the assumptions that  $\lambda_n/\sqrt{n} \rightarrow 0$  and  $\lambda_n n^{(\gamma-1)/2} \rightarrow \infty$ , Theorem 2 (Zou, 2006) states consistent variable selection

$$\lim_n P(\mathcal{A}_n^* = \mathcal{A}) = 1, \tag{27}$$

### A.3 Pseudo-Code: Online CPD

---

**Algorithm 1** `detect.CP` - verbal explanation in 2.7

---

- 1: # Last observation at  $t = T, h = 52, h_1 = 40, h_2 = 12$
  - 2: Run `estimate.AGFL` for  $t = [T-h+1, \dots, T-1]$
  - 3: Calculate one-step-ahead  $\hat{y}_{i,T}$  and  $e_{i,T}$  based on  $\hat{\alpha}_{i,J} = \hat{\beta}_{i,T-1}$
  - 4: Calculate  $\hat{y}_T$  and  $e_t$  by aggregating  $\hat{y}_{i,T}$  acc. to (2)
  - 5: Obtain  $(\lambda^*, \kappa^*)$  by running `run.ATD` for a grid of parameter combinations  $(\lambda, \kappa)$  based on the tuning sample  $t = T - h + 1, \dots, T - h + h_1$
  - 6: Calculate CP indicator  $cp.mac_T^{ARA}$  by `run.ATD` for the monitoring period  $t = [T - h + h_1 + 1, \dots, T]$  using  $(\lambda^*, \kappa^*)$  and report upswing/downswing/no CP at day  $T$
  - 7: Calculate evaluation measures (15)-(18)
  - 8: Optionally calculate DTA measures by `get.DTA.measures`
-

---

**Algorithm 2** estimate.AGFL

---

```
1: # At any time point  $t = T$ :
2: # Input:  $y_i \in \mathbb{R}^{T \times 1}, X_i \in \mathbb{R}^{T \times K}$  for each county  $i$ 
3: for each county  $i = 1, \dots, N$  do
4:   Estimate  $\hat{\beta}_{i,t}$  by solving (5) (with FISTA) for observations  $t = T - h + 1, \dots, T - 1$ 
5:   for  $t = T-h+1, \dots, T-1$  do
6:      $\hat{y}_{i,t+1} = X_{i,t+1} \cdot \hat{\beta}_{i,t}$  # One-step-ahead predictions assuming constant  $\hat{\beta}_{i,t}$ 
7:      $e_{i,t+1} = y_{i,t+1} - \hat{y}_{i,t+1}$  # One-step-ahead residuals
8:     # Offline CPD
9:      $cp.mic_{i,t+1}^{AGFL} = 0$ 
10:    if  $\hat{\beta}_{i,t+1} \neq \hat{\beta}_{i,t}$  then
11:       $cp.mic_{i,t+1}^{AGFL} = 1 \cdot \text{sgn}(e_{i,t+1})$  # Obtain reference CPs
12:    end if
13:  end for
14: end for
15: for  $t = T-h+1, \dots, T-1$  do
16:   # Count micro-level CPs
17:    $upswing.count_t^{AGFL} = \sum_{i=1}^N 1(cp.mic_{i,t+1}^{AGFL} = 1)$ 
18:    $downswing.count_t^{AGFL} = \sum_{i=1}^N 1(cp.mic_{i,t+1}^{AGFL} = -1)$ 
19:   # Output 1: Aggregated macro-level residuals, see Eq.(2)
20:    $e_{t+1} = \sum_{i=1}^N \omega_i \cdot e_{i,t+1}$ 
21: end for
22: # Output 2: Macro-level reference CPs
23: for  $t = T-h+1, \dots, T-1$  do
24:    $cp.mac_t^{AGFL} = 1$ , if  $upswing.count_t^{AGFL} \geq q.upswing^{AGFL}(0.05)$ 
25:    $cp.mac_t^{AGFL} = -1$ , if  $downswing.count_t^{AGFL} \geq q.downswing^{AGFL}(0.05)$ 
26: end for
```

---

---

**Algorithm 3** run.ATD

---

```
1: # Input:  $e_t$  from estimate.AGFL
2: # Parameters:  $\lambda, \kappa$ 
3: Set  $M_0 = 0$ 
4: for each period  $t = T - h + 1, \dots, T$  do
5:    $\hat{\sigma}_{t-1} = \frac{1}{t-T+h} \sum_{s=1}^{t-1} (e_s - \bar{e}_s)^2$  # Sample residual variance
6:    $M_t = \lambda M_{t-1} + (1 - \lambda)(e_t / \hat{\sigma}_{t-1})$  # Alarm statistic
7:   Set  $cp.mac_t^{ARA} = 0$  # CP Indicator variable
8:   # Decision rule (10)
9:   if  $M_{t-1} \leq \kappa$  then
10:    Set  $cp.mac_t^{ARA} = 1$ 
11:   end if
12:   if  $M_{t-1} > -\kappa$  then
13:    Set  $cp.mac_t^{ARA} = -1$ 
14:   end if
15: end for
16: # Output: Indicator variable  $cp.mac_t^{ARA}$ 
```

---

---

**Algorithm 4** get.DTA.measures
 

---

```

1: # Input:  $e_{i,t}$  from estimate.AGFL
2: # Parameters:  $\lambda, \kappa$ 
3: Set  $M_{i,0} = 0$ 
4: for each county  $i = 1, \dots, N$  do
5:   for each period  $t = T - h + 1, \dots, T$  do
6:      $\hat{\sigma}_{i,t-1} = \frac{1}{t-T+h} \sum_{s=1}^{t-1} (e_{i,s} - \bar{e}_i)^2$  # Sample residual variance
7:      $M_{i,t} = \lambda M_{i,t-1} + (1 - \lambda)(e_{i,t}/\hat{\sigma}_{i,t-1})$  # Alarm statistic
8:     Set  $cp.mic_{i,t}^{DTA} = 0$  # CP Indicator variable
9:     # Decision rule (10)
10:    if  $M_{i,t-1} \leq \kappa$  then
11:      Set  $cp.mic_{i,t}^{DTA} = 1$ 
12:    end if
13:    if  $M_{i,t-1} > -\kappa$  then
14:      Set  $cp.mic_{i,t}^{DTA} = -1$ 
15:    end if
16:  end for
17: end for
18: # Calculate homogeneity measures
19:  $P_t^d = \sum_{i=1}^N \omega_i \cdot \mathbb{1}(cp.mic_{i,t}^{DTA} = d), d \in \{1, -1, 0\}$ 
20: Obtain  $cp.share_t, cp.excess_t^+, cp.homogeneity_t$  (Eq. (12)-(14))
21: # Output: Indicator variable  $cp.mic_{i,t}^{DTA}$  and homogeneity measures (12)-(14)

```

---

#### A.4 Class-specific 2x2 Confusion Matrices

**Upswings:**

Predicted/Actual	Upswing	No Upswing
Upswing	$TU$	$FU^- + FU^0$
No Upswing	$FN^+ + FD^+$	$TN + TD + FD^0 + FN^-$

Table 3: Confusion matrix for upswings

**Neutrals:**

Predicted/Actual	Neutral	No Neutral
Neutral	$TN$	$FN^+ + FN^-$
No Neutral	$FU^0 + FD^0$	$TU + TD + FD^+ + FU^-$

Table 4: Confusion matrix for neutrals

**Downswings:**

<b>Predicted/Actual</b>	Downswing	No Downswing
Downswing	$TD$	$FD^+ + FD^0$
No Downswing	$FU^- + FN^-$	$TU + TN + FN^+ + FU^0$

Table 5: Confusion matrix for downswings

### A.5 Data Set - Dependent Variable on County Level (*Landkreise*)

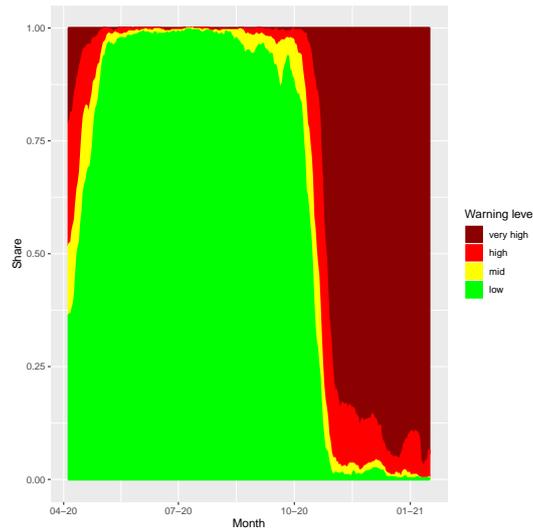


Figure 3: According to the German warning level system, the 401 German counties are classified into category low ( $y_t < 35$ ), mid ( $35 \leq y_t < 50$ ), high ( $50 \leq y_t < 100$ ) and very high ( $y_t > 100$ ). Plotting the warning level distribution over time depicts an illustrative indicator for the pandemic development.

### A.5.1 Dependent variable

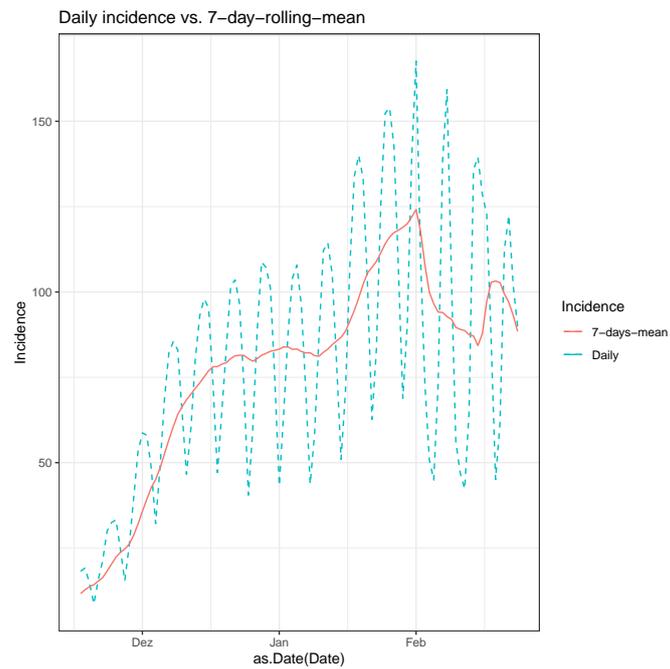


Figure 4: Time series of the incidence (dashed line) and its 7-days moving average (solid line) for whole Germany. The weekly reporting pattern is smoothed by the 7-days-averaging.

## A.5.2 Dependent variable - Micro-, Mezzo- and Macro-level

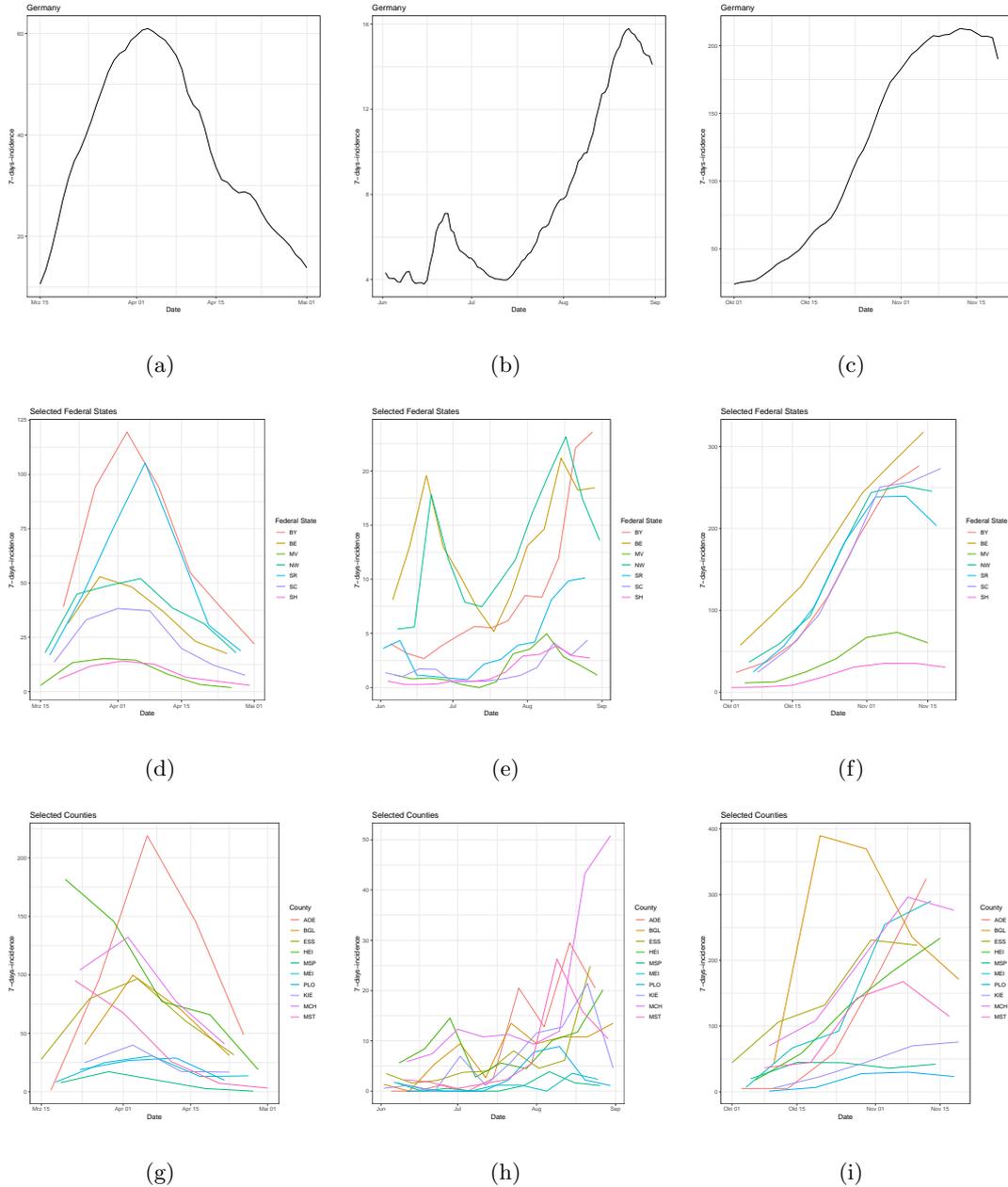


Figure 5: Time series of incidence (7-days moving average) for whole (a-c)) Germany, (d-f)) selected federal state and (g-i)) selected counties.

Periods: '2020-03-15'-'2020-05-01' (a, d, g)), '2020-06-01'-'2020-09-01' (b, e, h)), '2020-10-01'-'2020-11-19' (c, f, i)) .

When looking at the macro-level (Figures 5(a), 5(b), 5(c)), there is only an aggregated line for all sub-units which contains no information about the heterogeneity in the sub-units. Having a look at the mezzo-level (Figures 5(d), 5(e), 5(f)), the incidence across different federal states

exhibits a high degree of diversity. E.g. Figure 5(e) reveals that the peak in late June visible in 5(b) boils down to only few federal states *Nordrhein-Westfalia* and *Bavaria* while other federal states had a constant incidence. A dispersed behaviour of the federal states is also observed in November when some federal states faced decreasing incidence (e.g. *Mecklenburg Western Pomerania*, *Nordrhein-Westfalia*) while other federal state still experienced a further increase of the incidence (e.g. *Saxony* or *Berlin*). Naturally, the micro-level (Figures 5(g), 5(h), 5(i)) provides the most heterogeneous picture. Particularly during the first (Figure 5(g)) and the second (Figure 5(i)) wave, the different timing and strength of the infection waves becomes evident which reflects the variety of regional events and policies.

### A.5.3 Mobility data

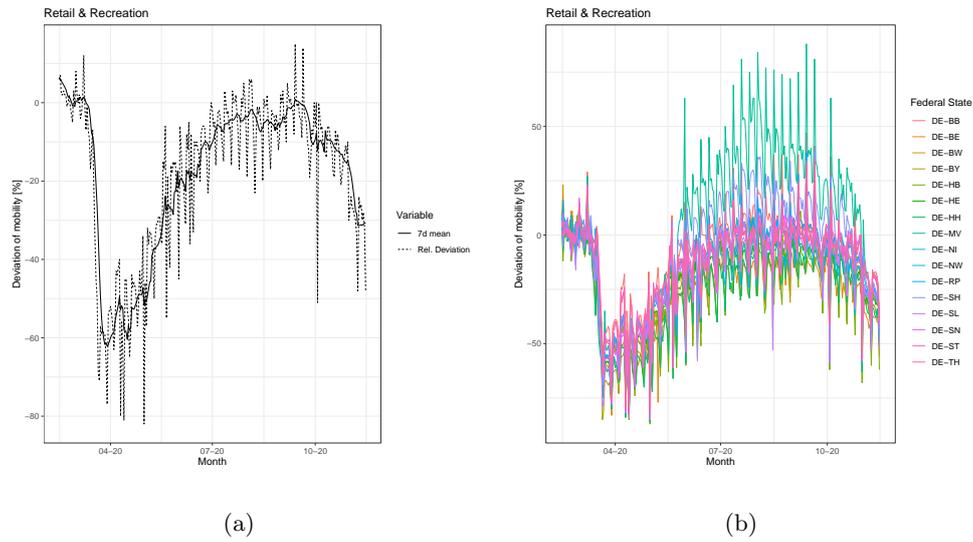


Figure 6: Mobility deviation against the reference period in the resort 'Retail and Recreation' (i.e. activity of the population in museums, cinemas, cafés, restaurants, shopping zones etc.) for (a) whole Germany and (b) all single German federal states.

## A.6 Benchmark CPs

### A.6.1 Micro-level CPs

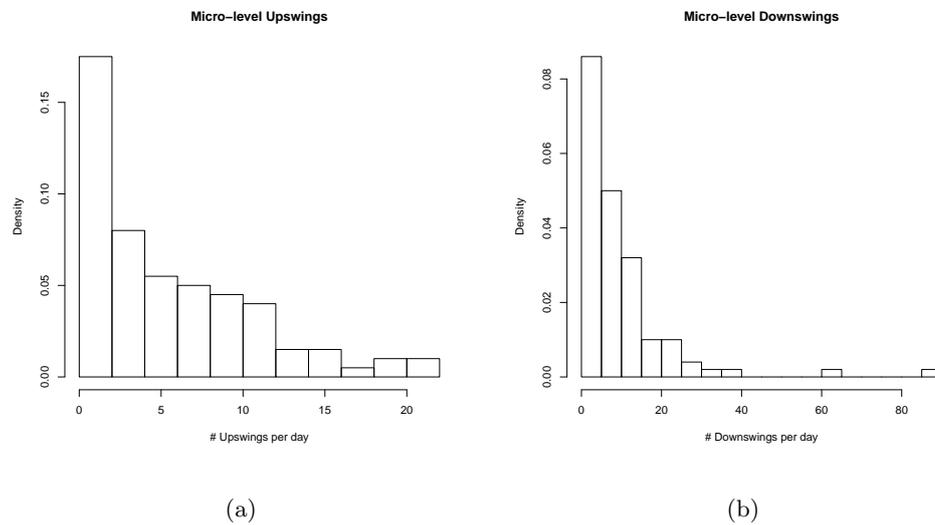


Figure 7: The histograms of (a) upswings and (b) downswings reveal that micro-level downswings tend to occur synchronously while micro-level upswings arise dispersed over time.

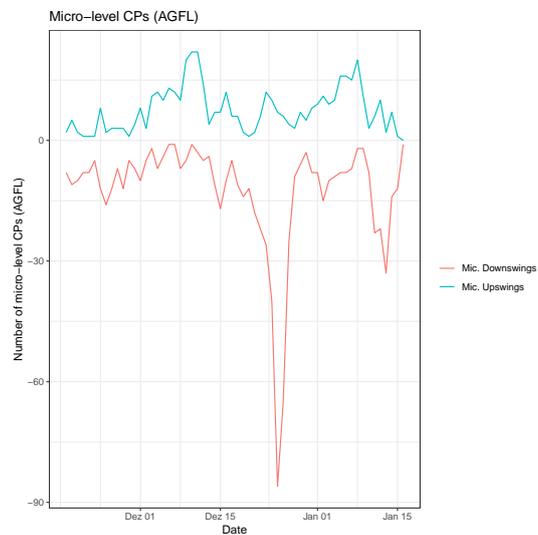
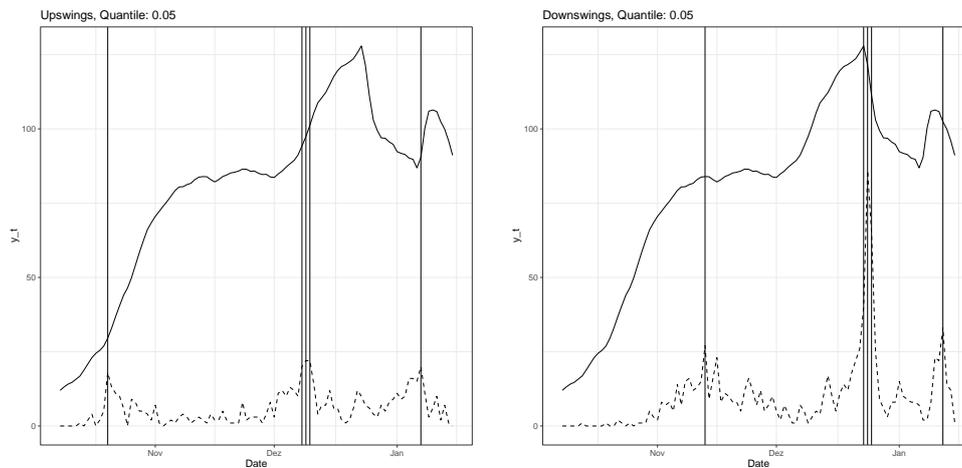


Figure 8: Number of micro-level reference up- and downswings by AGFL (5) during the period from '2020-11-18' until '2021-01-16'.

## A.6.2 Macro-level CPs



(a) Macro-level Upswings (AGFL)

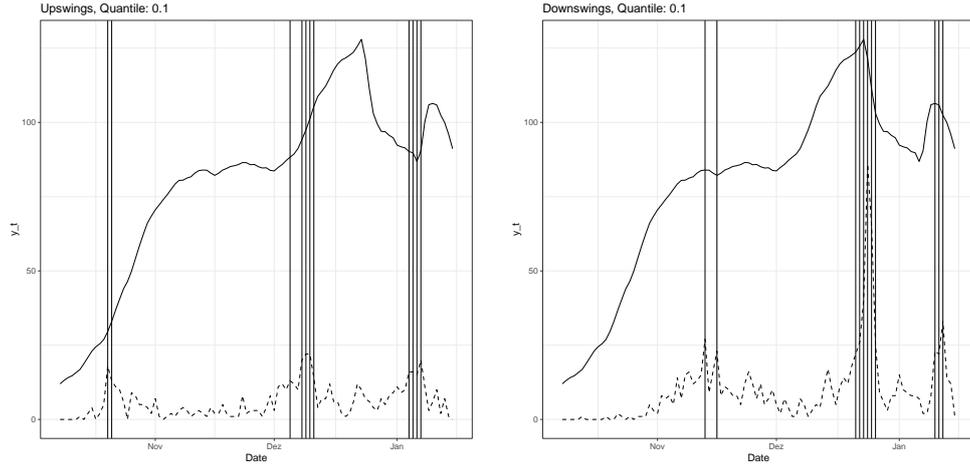
(b) Macro-level Downswings (AGFL)

Figure 9: The graphs depict the macro-level reference upswings (a) and downswings (b) (vertical lines) defined by the upper 5% of the number of micro-level CPs (dashed line) and the dependent variable  $y_t$  (solid line, (20)).

Upswings: '2020-12-07', '2020-12-08', '2020-12-09', '2021-01-06'

Downswings: '2020-12-22', '2020-12-22', '2020-12-23', '2020-12-24'

Table 6: Reference macro-level CPs (AGFL) based on the 5% quantile.



(a) Macro-level Upswings (AGFL)

(b) Macro-level Downswings (AGFL)

Figure 10: The graphs depict the macro-level reference upswings (a) and downswings (b) (vertical lines) defined by the upper 10% of the number of micro-level CPs (dashed line) and the dependent variable  $y_t$  (solid line, (20)).

Upswings: '2020-12-07', '2020-12-08', '2020-12-09', '2021-01-03', '2021-01-04', '2021-01-06'

Downswings: '2020-12-21', '2020-12-22', '2020-12-23', '2020-12-24', '2020-12-25', '2021-01-11'

Table 7: Reference macro-level CPs (AGFL) based on the 10% quantile.

## A.7 Online CPD

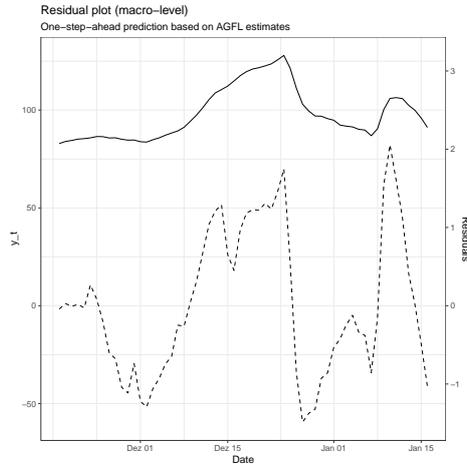


Figure 11: Aggregated residuals (dashed line) based on the one-step-ahead prediction from (5) and the dependent variable  $y_t$  (solid line, (20)). [The residual plot provides an overview over the occurrence of CPs and their CP direction. According to the development of the incidence, the aggregated one-step-ahead residuals turn positive during upswing periods and negative in downswing periods. In the given example, the residuals instantaneously become positive in response to the upswings in the mid of December 2020. The residuals turn negative during the temporary downswing of the COVID-19 infections after Christmas 2020.]

### A.7.1 Macro-level CPs reported by ARA

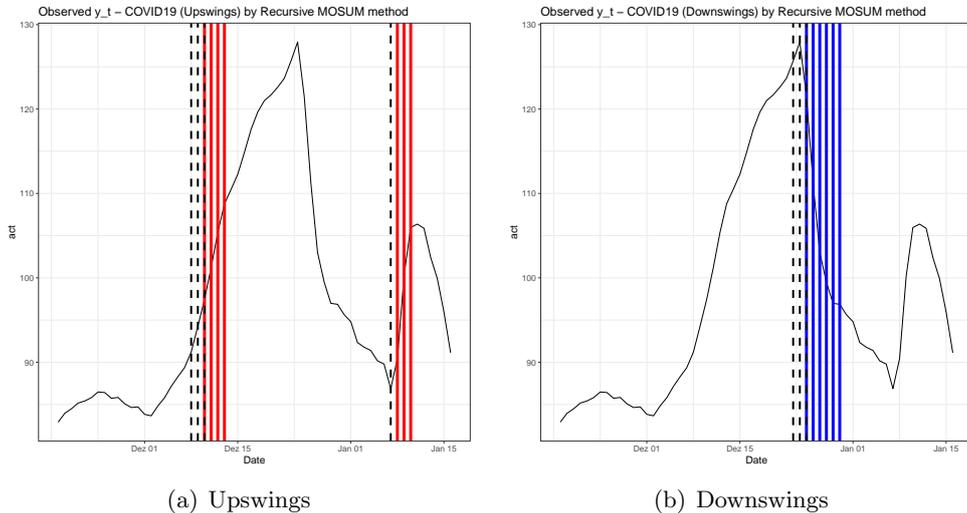


Figure 12: Reported (a) upswings and (b) downswings based on model ARA using a monitoring window length of 10 days together with the incidence (solid line). The reference change points (5% quantile) are marked as dashed vertical lines respectively, whereas the coloured vertical lines represent the up- and downswings detected by ARA.

Optimal parameters:  $\kappa^* = 1.5$ ,  $\lambda^* = 0.74$ ,  $\gamma_1^* = 9.11$ ,  $\gamma_2^* = -5.92$

Reported CPs:

- $\kappa = 1$

- **Upswings:** '2020-12-08', '2020-12-09', '2020-12-10', '2020-12-11', '2020-12-12', '2020-12-13', '2021-01-02', '2021-01-03', '2021-01-08', '2021-01-09', '2021-01-10', '2021-01-11'
- **Downswings:** '2020-11-27', '2020-11-28', '2020-11-30', '2020-12-01', '2020-12-25', '2020-12-26', '2020-12-27', '2020-12-28', '2020-12-29', '2020-12-30', '2020-12-31'

- $\kappa = 1.5$

- **Upswings:** '2020-12-10', '2020-12-11', '2020-12-12', '2020-12-13', '2021-01-08', '2021-01-09', '2021-01-10'
- **Downswings:** '2020-12-25', '2020-12-26', '2020-12-27', '2020-12-28', '2020-12-29', '2020-12-30'

- $\kappa = 2$

- **Upswings:** '2021-01-08', '2021-01-09'
- **Downswings:** '2020-12-25', '2020-12-26', '2020-12-27', '2020-12-28', '2020-12-29', '2020-12-30'

## A.7.2 CPTT

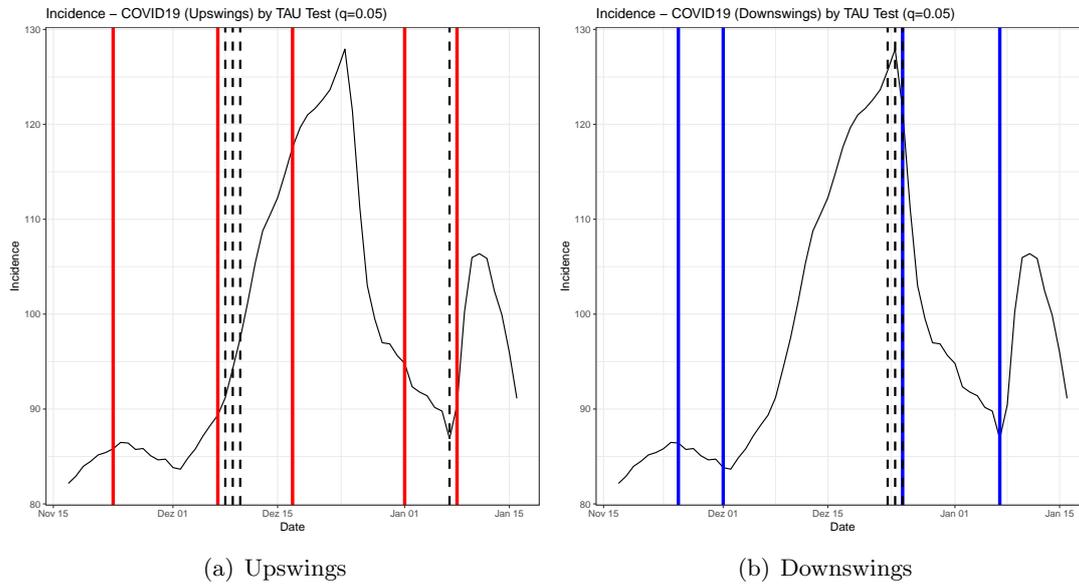


Figure 13: Reported (a) upswings and (b) downswings based on CPTT using a monitoring window length of 10 days together with the incidence (solid line). The reference change points (5% quantile) are marked as dashed vertical lines respectively, whereas the colored vertical lines represent the up- and downswings detected by CPTT.

- **Upswings:** '2020-12-16', '2020-12-23', '2021-01-08', '2021-01-09'
- **Downswings:** '2020-11-30', '2020-12-24', '2020-12-25', '2020-12-26', '2021-01-15'

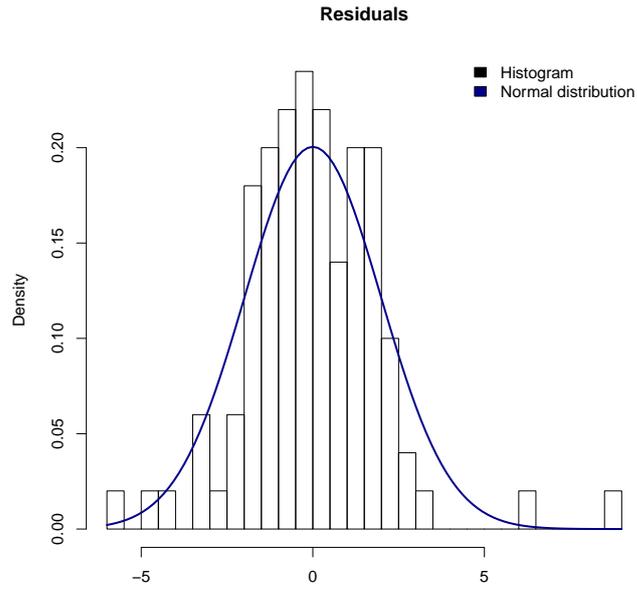


Figure 14: According to the JB test results (2), the skewness of the empirical distribution of the residuals is not symmetric due to some outliers and a slightly left-shifted peak. Moreover, the kurtosis seems to fit better to the normal distribution, but is still rejected by the JB test. The comparably broad peak and the comparably strong tails are non-normal.

### A.7.3 MORA

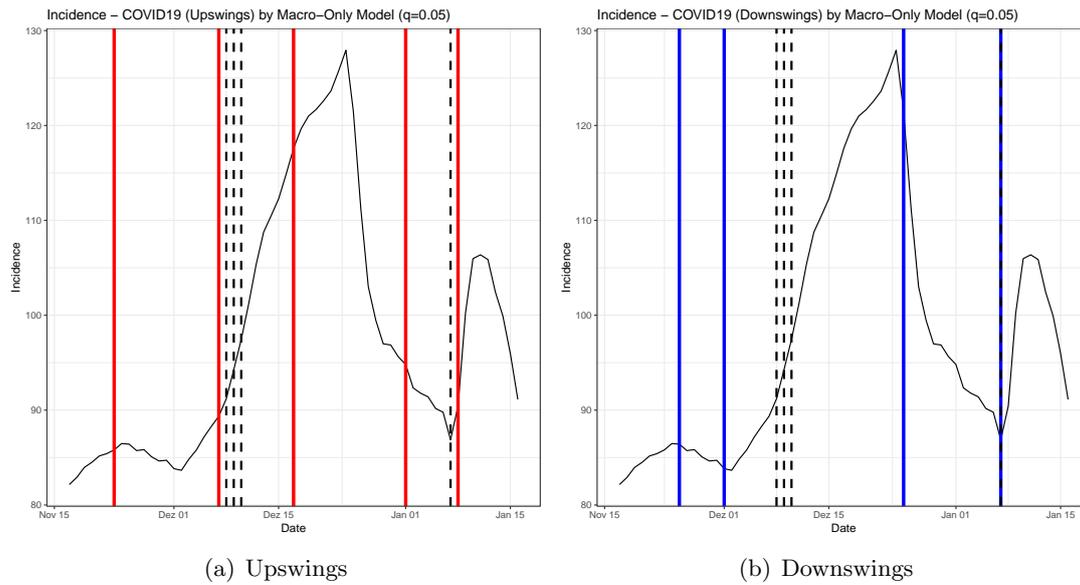


Figure 15: Reported (a) upswings and (b) downswings based on MORA using a monitoring window length of 10 days together with the incidence (solid line). The reference change points (5% quantile) are marked as dashed vertical lines respectively, whereas the colored vertical lines represent the up- and downswings detected by the MORA.

- **Upswings:** '2020-11-22', '2020-12-06', '2020-12-16', '2020-12-31', '2021-01-07'
- **Downswings:** '2020-11-24', '2020-11-30', '2020-12-24', '2021-01-06'

## A.8 Online CPD: Model Comparison

### A.8.1 Reference CPs: 5% quantile based on AGFL

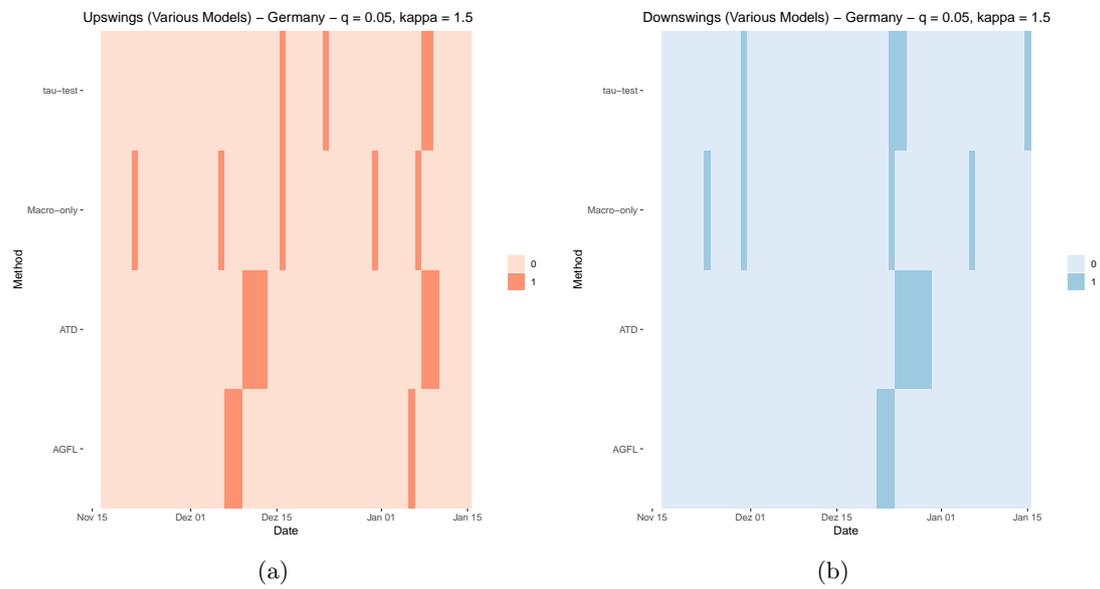


Figure 16: The figures oppose the reported CPs of the ARA, the MORA and the CPTT with the reference change points based on the AGFL.

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	1	6	0
Pred. Neutral	3	43	2
Pred. Downswing	0	4	1

Table 8: Confusion matrix ARA (macro)

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	72	2681	54
Pred. Neutral	343	18349	556
Pred. Downswing	39	1830	136

Table 9: Confusion matrix ARA (micro)

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	0	3	1
Pred. Neutral	4	47	0
Pred. Downswing	0	3	2

Table 10: Confusion matrix CPTT (macro)

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	30	1519	55
Pred. Neutral	399	19548	504
Pred. Downswing	25	1793	187

Table 11: Confusion matrix CPTT (micro)

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	0	5	0
Pred. Neutral	3	46	2
Pred. Downswing	1	2	1

Table 12: Confusion matrix MORA (macro)

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	46	1912	47
Pred. Neutral	376	19460	615
Pred. Downswing	32	1488	84

Table 13: Confusion matrix MORA (micro)

	Upswings	Neutrals	Downswings
AGFL	0.07	0.88	0.05
ARA	0.12	0.80	0.08
CPTT	0.07	0.85	0.08
MORA	0.08	0.85	0.07

Table 14: The prediction rates of the three classes by online CPD model.

	Acc. Macro	Acc. Micro
ARA	0.75	0.77
CPTT	<b>0.82</b>	<b>0.82</b>
MORA	0.78	0.81

Table 15: Prediction accuracy by online CPD model.

	TPR mac.	TPR mic.	FPR mac.	FPR mic.
ARA	<b>0.25</b>	<b>0.16</b>	<b>0.43</b>	0.51
CPTT	0.00	0.07	0.44	<b>0.39</b>
MORA	0.00	0.10	0.50	0.47

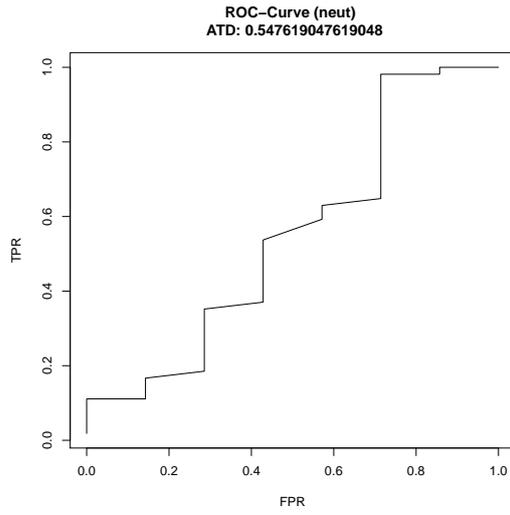
Table 16: Comparison - TPR vs. FPR on macro- and micro-level w.r.t. class 'Upswings' by online CPD method.

	TPR mac.	TPR mic.	FPR mac.	FPR mic.
ARA	0.81	0.80	0.71	<b>0.75</b>
CPTT	<b>0.89</b>	<b>0.86</b>	<b>0.57</b>	<b>0.75</b>
MORA	0.87	0.85	0.71	0.83

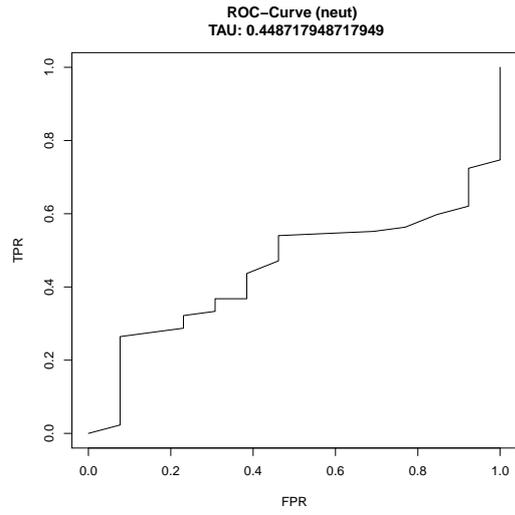
Table 17: Comparison - TPR vs. FPR on macro- and micro-level w.r.t. class 'Neutrals' by online CPD method.

	TPR mac.	TPR mic.	FPR mac.	FPR mic.
ARA	0.33	0.18	<b>0.07</b>	0.08
CPTT	<b>0.67</b>	<b>0.25</b>	0.05	0.08
MORA	0.33	0.11	0.05	<b>0.07</b>

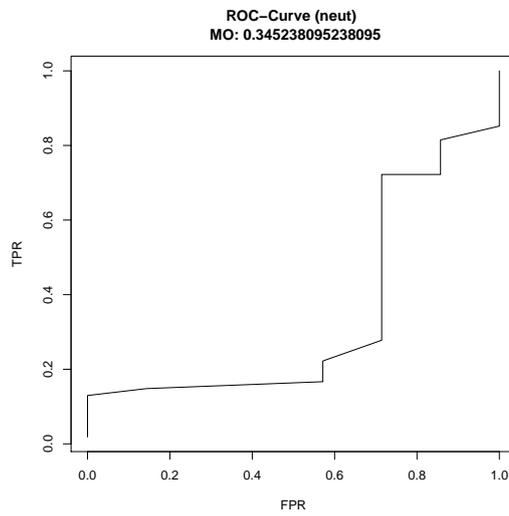
Table 18: Comparison - TPR vs. FPR on macro- and micro-level w.r.t. class 'Downswings' by online CPD method.



(a) ARA



(b) CPTT



(c) MORA

Figure 17: AUC scores achieved by various online CPD models

### A.8.2 Reference CPs (10% quantile):

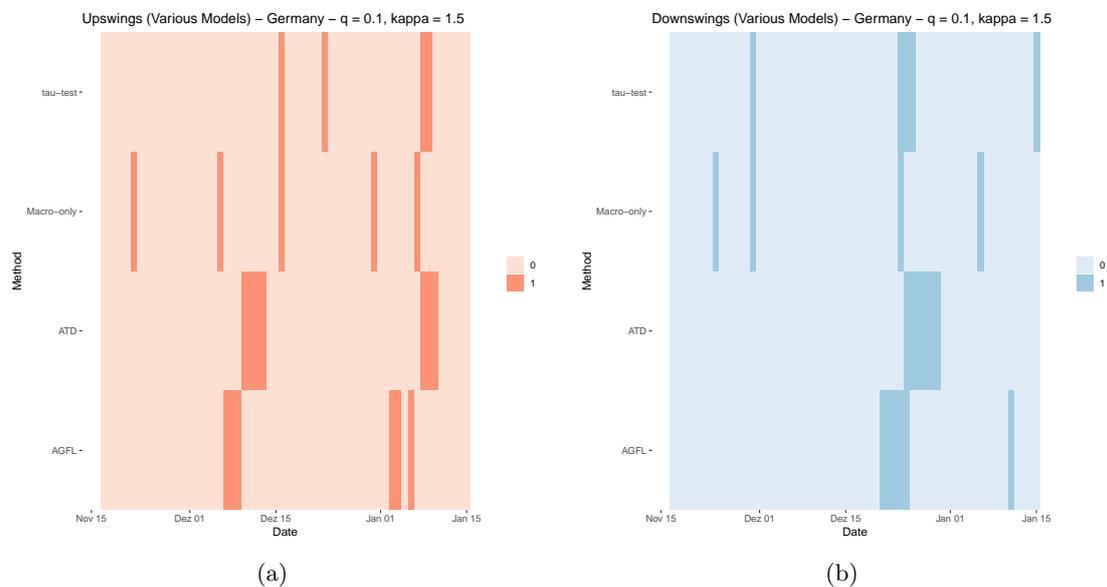


Figure 18: The figures oppose the reported CPs of the ARA and of the CPTT with the reference change points based on the AGFL. The CPTT fails to detect the upswing around 10<sup>th</sup> of December 2020. In contrast, the CPTT identifies the downswing at 12<sup>th</sup> of January 2021 with a small delay of 3 days.

	Upswings	Neutrals	Downswings
AGFL	0.10	0.80	0.10
ARA	0.12	0.80	0.08
CPTT	0.07	0.85	0.08
MORA	0.08	0.85	0.07

Table 19: The prediction rate of the three classes by online CPD model.

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	1	5	1
Pred. Neutral	5	39	4
Pred. Downswing	0	4	1

Table 20: Confusion matrix ARA (macro).

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	72	2681	54
Pred. Neutral	343	18349	556
Pred. Downswing	39	1830	136

Table 21: Confusion matrix ARA (micro).

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	0	3	1
Pred. Neutral	6	43	2
Pred. Downswing	0	2	3

Table 22: Confusion matrix CPTT (macro).

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	30	1519	55
Pred. Neutral	399	19548	504
Pred. Downswing	25	1793	187

Table 23: Confusion matrix CPTT (micro).

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	0	5	0
Pred. Neutral	5	41	5
Pred. Downswing	1	2	1

Table 24: Confusion matrix MORA (macro).

	Act. Upswing	Act. Neutral	Act. Downswing
Pred. Upswing	46	1912	47
Pred. Neutral	376	19460	615
Pred. Downswing	32	1488	84

Table 25: Confusion matrix MORA (micro).

	Acc. Macro	Acc. Micro
ARA	0.68	0.77
CPTT	<b>0.77</b>	<b>0.82</b>
MORA	0.70	0.81

Table 26: Prediction accuracy (micro- and macro-level) by online CPD model.

	TPR mac.	TPR mic.	FPR mac.	FPR mic.
ARA	0.17	0.16	0.38	0.51
CPTT	0.00	0.07	0.36	0.39
MORA	0.00	0.10	0.38	0.47

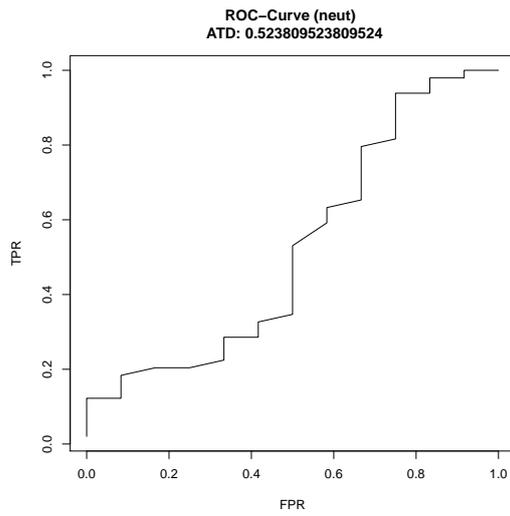
Table 27: TPR vs. FPR on macro- and micro-level w.r.t. class 'Upswings' by online CPD method.

	TPR mac.	TPR mic.	FPR mac.	FPR mic.
ARA	0.81	0.80	0.75	0.75
CPTT	0.90	0.86	0.67	0.75
MORA	0.85	0.85	0.83	0.83

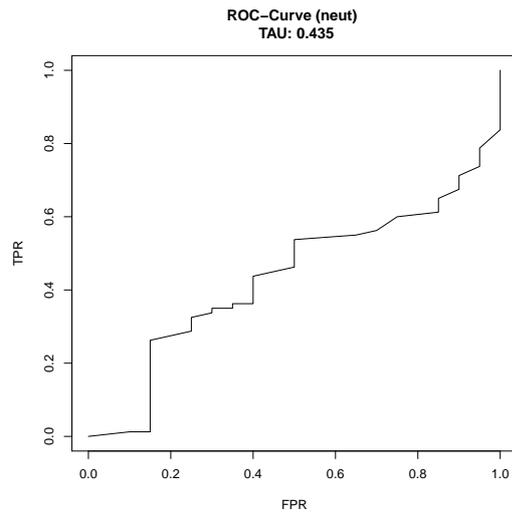
Table 28: TPR vs. FPR on macro- and micro-level w.r.t. class 'Neutrals' by online CPD method.

	TPR mac.	TPR mic.	FPR mac.	FPR mic.
ARA	0.17	0.18	0.07	0.08
CPTT	0.50	0.25	0.04	0.08
MORA	0.17	0.11	0.06	0.07

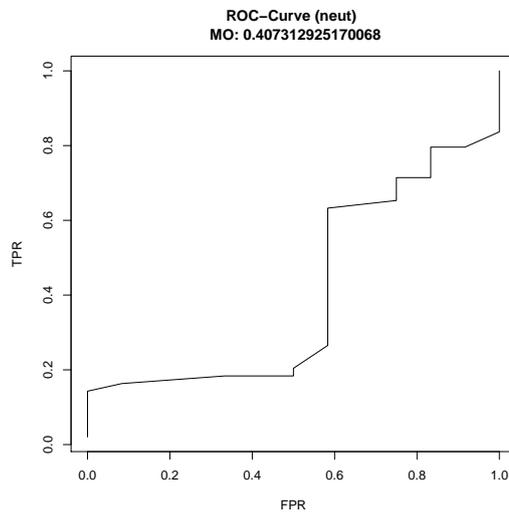
Table 29: TPR vs. FPR on macro- and micro-level w.r.t. class 'Downswings' by online CPD method.



(a) ARA



(b) CPTT



(c) MORA

Figure 19: AUC scores achieved by various online CPD models

### A.8.3 Comparison ARA with lower sensitivity ( $\kappa = 2$ )

	$q = 0.95$		$q = 0.90$	
	$\kappa = 1.5$	$\kappa = 2$	$\kappa = 1.5$	$\kappa = 2$
ARA	0.733	0.783	0.700	0.733
CPTT	<b>0.817</b>	<b>0.817</b>	<b>0.767</b>	<b>0.767</b>
MORA	0.783	0.783	0.633	0.633

Table 30: (Macro) prediction accuracy for the different CPD models for different threshold values of ARA and different reference CP definitions.

## A.9 Regional Analysis

### A.9.1 Regions

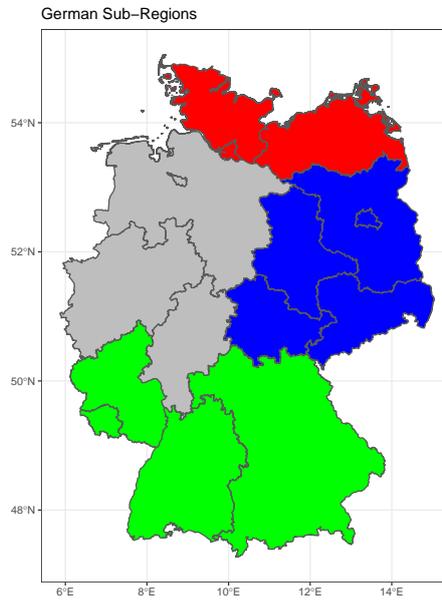


Figure 20: Regions

**West** (grey): Nordrhein-Westfalia, Lower Saxony, Bremen, Hesse

**East** (blue): Thuringia, Saxony, Saxony-Anhalt, Brandenburg, Berlin

**South** (green): Baden-Württemberg, Saarland, Rhineland-Palatinate, Bavaria

**North** (red): Schleswig-Holstein, Mecklenburg-Vorpommern, Hamburg

### A.9.2 CPs by Region (ARA vs. AGFL) - reference: 5% quantile AGFL

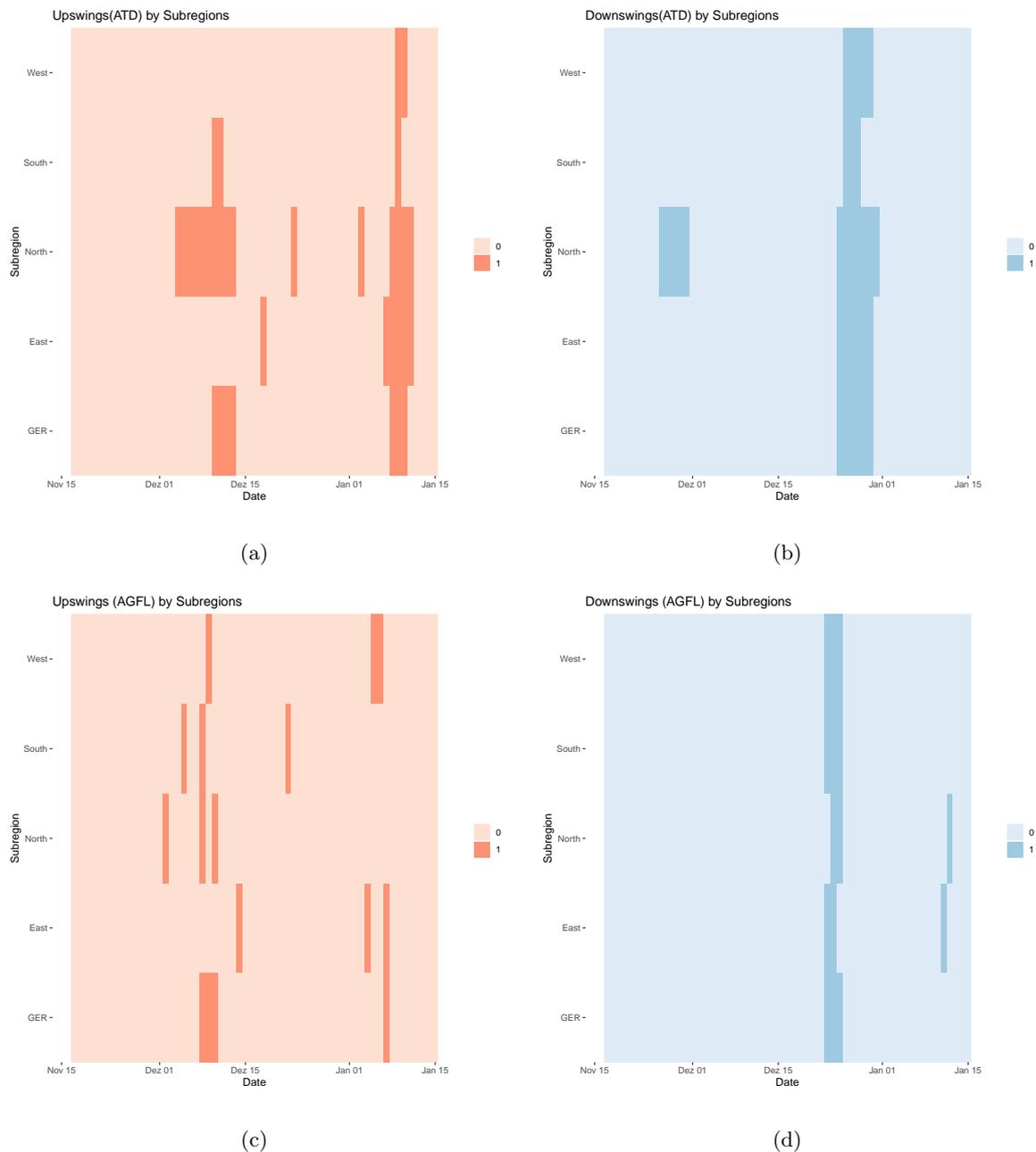


Figure 21: In (a) and (b), the CPs detected by the real-time model ARA are marked for the considered sample period '2020-11-16' - '2021-01-15'. These CPs are opposed to the reference CPs identified by AGFL ((c) and (d)). [The upswings are detected fairly well on the country-level (5th row). The downswings are detected in time, but also false alarms are reported in the first half of the observation period. On the regional level, the results are worse. Some true CPs are not reported and furthermore, there are numerous false alarms.]

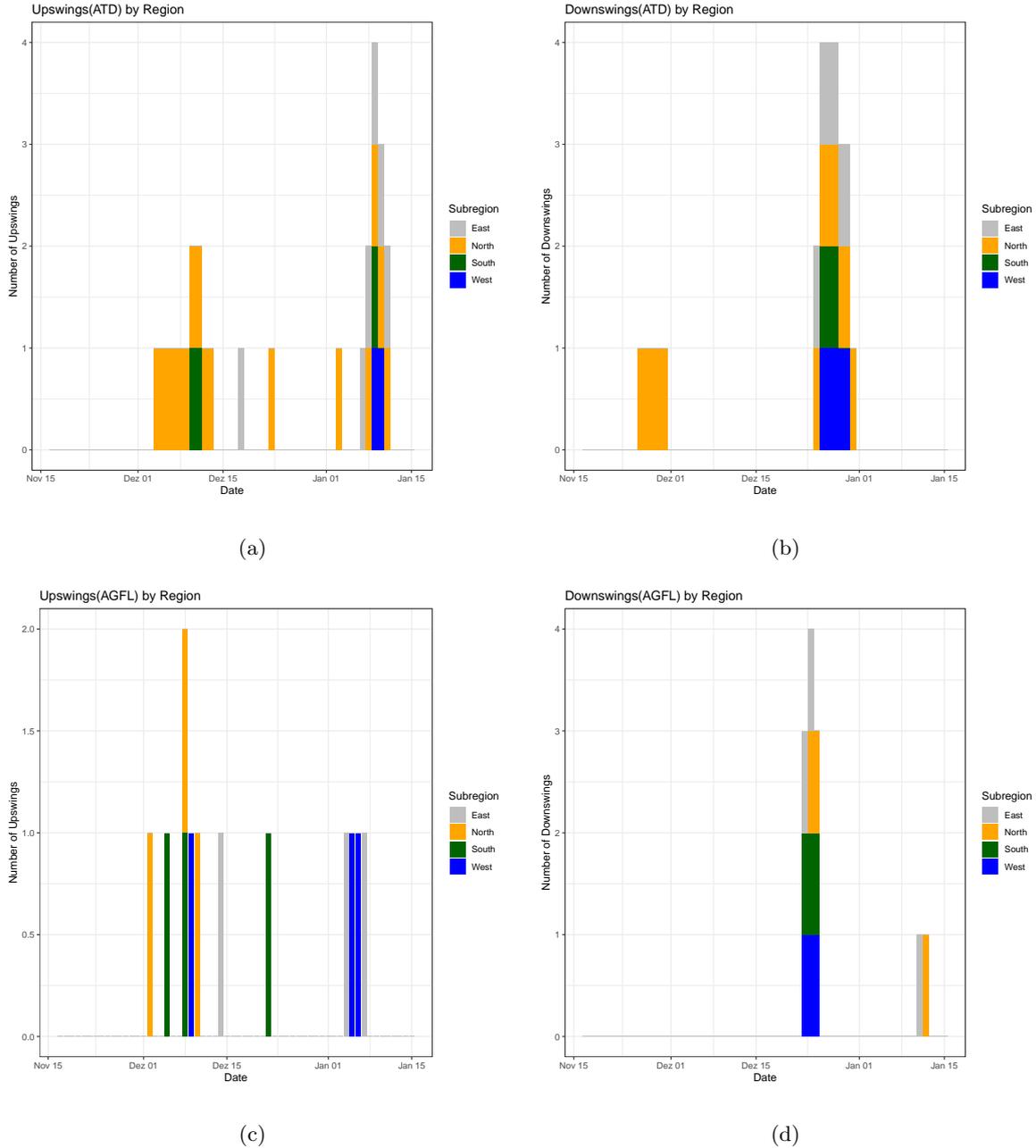


Figure 22: In (a) and (b), the CPs detected by the real-time model ARA are marked for the considered sample period '2020-11-16' - '2021-01-15'. These CPs are opposed to the reference CPs identified by AGFL ((c) and (d)). [The figure shows reported CPs by the ARA on the regional level. The downswings are detected in time, but also false alarms are reported in the first half of the observation period. On the regional level, the ARA detection performance is only partially good. Some reference CPs are not reported in all regions (e.g. regional upswings in all regions around 10<sup>th</sup> of December) and furthermore, there are few false alarms (e.g. ARA downswings in the beginning of December).]

### A.9.3 CPs by Region (ARA vs. AGFL) - reference: 10% quantile AGFL

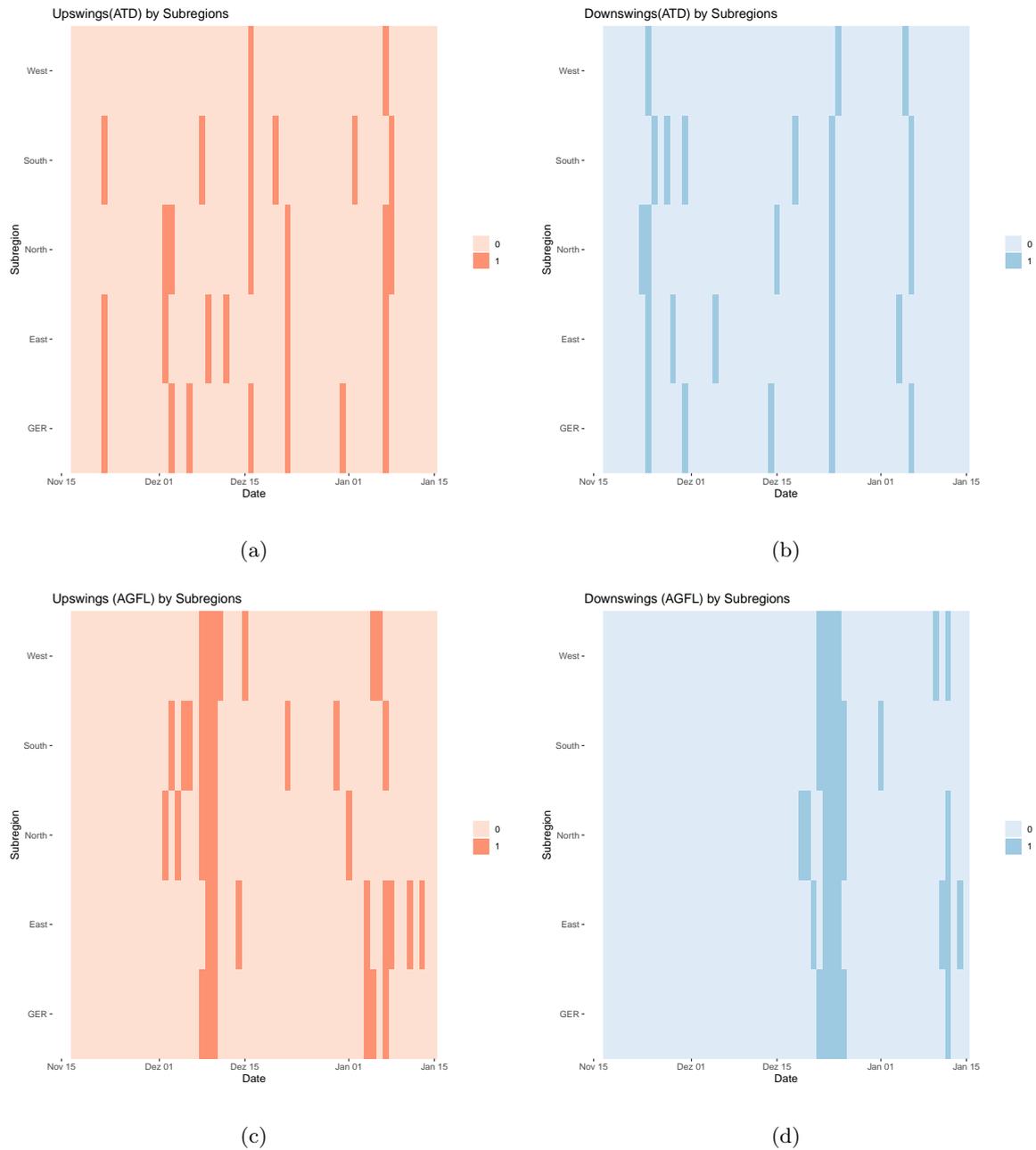
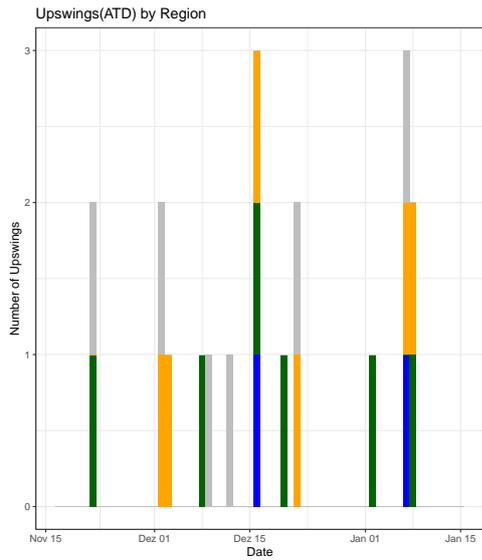
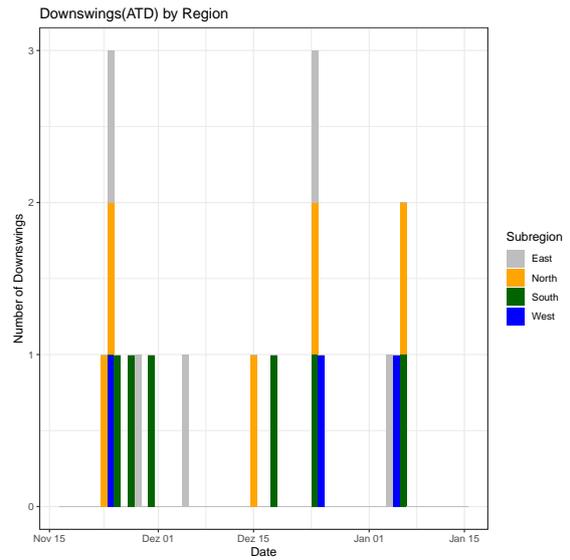


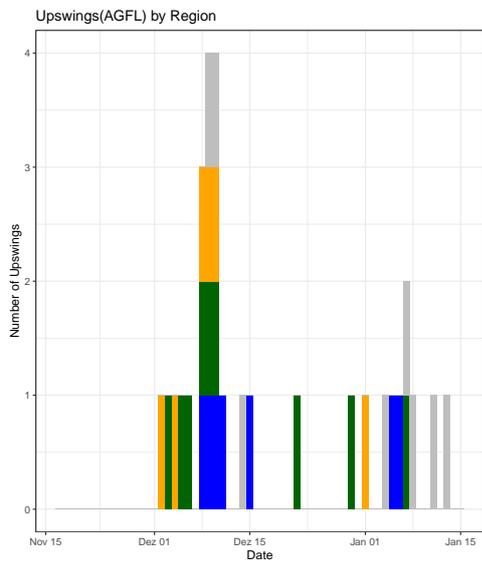
Figure 23: In (a) and (b), the CPs detected by the real-time model ARA are marked for the considered sample period '2020-11-16' - '2021-01-15'. These CPs are opposed to the reference CPs identified by AGFL ((c) and (d)).



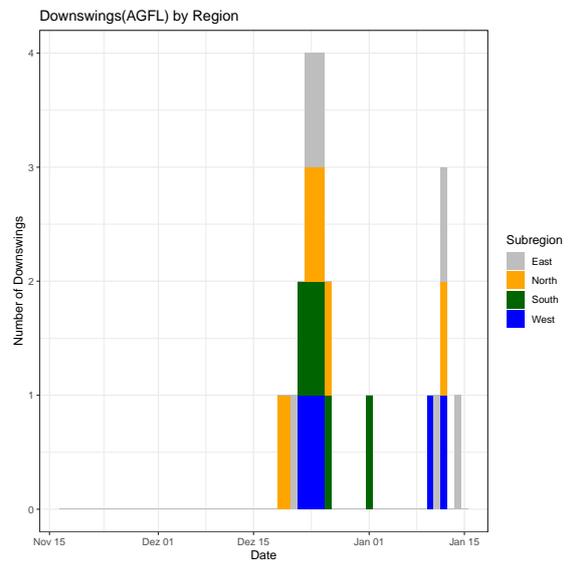
(a)



(b)



(c)



(d)

Figure 24: In (a) and (b), the CPs detected by the real-time model ARA are marked for the considered sample period '2020-11-16' - '2021-01-15'. These CPs are opposed to the reference CPs identified by AGFL ((c) and (d)).