MODELLING HUMAN GESTURES AND CONTROL BEHAVIOUR FROM MEASURED DATA

Roderick Murray-Smith

Department of Computing Science,
University of Glasgow,
Glasgow G12 8QQ,
Scotland, UK

Abstract: Approaches to modelling human control behaviour and gestures from measurements are reviewed. Multiple-model and nonparametric statistical models can be used to model and recognise human gestures, and control behaviour. Functional Data Analysis (FDA) techniques with nonparametric models, and regularisation, can be used to explore the data. The applicability of the approach is demonstrated in a gesture recognition test based on complex hand movements.

Keywords: Algorithms & methodology, human–machine interface, learning models of human control behaviour from experimental data, gesture recognition, functional data analysis.

1. INTRODUCTION

This paper describes approaches to the estimation of human states using empirically-derived behaviour models. This makes it feasible to recognise operators' goals and 'modes' of behaviour from their actions, as well as potentially allowing more natural gesture-driven interaction with control systems.

There are several goals:

- To develop approaches which could be used to estimate and predict human skills, such that we would be able to learn individual preferences and expectations, and detect characteristic features and types of human user.
- To develop models which can be used to improve human–machine coordination. If the intentions, goals and preferences, and the accompanying skills of the human operator were known, the human–machine interaction problem would be to coordinate, adapt, and configure the automatic control system to ensure satisfactory performance of the full human–machine control system. Unfortunately, the states of the human are usually not known – at best they can be estimated, using user models.
- To link modern nonparametric statistical modelling approaches with dynamic systems theory, applied to the challenging nonlinear identification tasks associated with models of human behaviour.
- The methods outlined here involve classification of segments of state trajectories – pattern recognition of time-series. Here we examine classification of measured user states, but the same tools might be useful in segmentation and pattern recognition of the controlled process, for either control, visualisation or warning purposes.

Because of the complexity of human behaviour, and the richness of sensing and state, no conceivable model will be able to predict exactly what the human will do. In this paper we will use a probabilistic framework for the representation of human behaviour. This provides a common framework for describing the uncertainty in both the human and technical aspects of our system and allows us to develop models which for the given task behave statistically as a human would, within a certain scope of interest.

2. HUMAN GESTURES AND CONTROL

Humans are capable of generating complex continuous motor actions which include natural movements such as walking, reaching, throwing etc, and explicit control of a task, or generation of complex trajectories for handwriting, or in gesturing for communication. These complex actions all combine open and closed loop control to varying degrees.

The handwriting and gesture modelling tasks have some common features: timing of execution can vary greatly, as can the final path drawn, so finding the right invariant features is challenging. Different measurement systems will have different types of error, depending on whether position or higher derivatives are measured directly, such as velocity or acceleration components. The model fitting process should be able to deal with measurement error robustly, which motivates the use of regularisation methods.

The motivation for modelling the behaviour of such motions in this paper is not to better understand the physiological nature of the generating mechanism in humans, but is more directed at use of empirical models in improving human-computer interaction. In pursuing this goal, it is also of interest to see how modern nonparametric statistical models, based directly on measurements rather than physically motivated models, can aid both in producing working interactive systems. The basic idea is to represent gestures and movements with multiple models, where each submodel represents some part of a gesture or movement, and where we piece these together to segment and recognise complex gestures or series of actions. The task of gesture modelling is a challenging test of the flexibility of such nonlinear model structures, and it is hoped that this will stimulate further development of the methodology.

2.1 *Handwriting models*

Handwriting is the result of a complex bio-mechanical system driven by a control process which generates the oscillatory behaviour needed.

Singer and Tishby (1994) show how a relatively simple model which switches between systems of coupled second order linear differential equations can generate surprisingly lifelike cursive handwriting using a low-rate modulation of parameters within a small range of discrete values. This can also be used for classification. The approach, although derived independently, could be linked to the switching multiple-model approach described in section 3.1. An example of output is shown in figure 1 (this is not particularly legible, but is intended to give a feel for how simple parameter variation in second-order systems could lead to cursive script. See (Singer and Tishby, 1994) for more impressive examples.

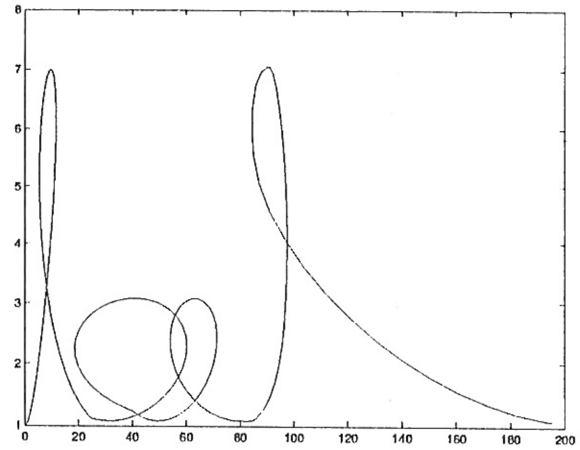


Fig. 1. An example of cursive-like script generated by arbitrary changes in the parameters of second order differential equations in x and y components of the pen velocity.

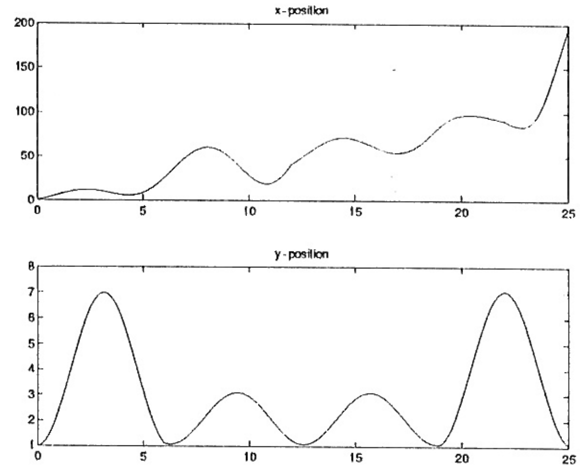


Fig. 2. The x and y time-series associated with the figure above.

In (Ramsay, 2000), functional data analysis techniques are used to analyse time-series generated by Chinese handwriting. These techniques use robust smoothing methods to estimate smoothly time-varying parameters of second order differential equations to describe the complex trajectories produced by humans. For general background on this approach, see (Ramsay and Silverman, 1997).

2.2 *Gesture models*

A related interesting area is that of human gesture recognition. Here we interpret the word *gesture* as being 'a motion or sequence of actions which has a special interpretation in a given context'. This could underpin novel approaches to human-machine interaction. The use of gestures is particularly interesting in environments in which keyboard use is infeasible, such as mobile applications. It also has interesting potential for disabled users.

Examples of related work include (Wilson and Bobick, 1995), and a general review can be found in (Harling and Edwards, 1997). Vision-based methods are becoming increasingly popular, e.g. (Bregler *et al.*, 1998; Bregler, 1999). Use of acceleration measurements for gesture recognition is described in (Sawada *et al.*, 1998). Ju *et al.* (2000) investigate the use of Electromyographic signals in finger gesture recognition.

3. MODELLING APPROACHES

We describe two general approaches to modelling human behaviour. One is to model the gesture as if it were composed by switching between simple submodels. Another is to generate smooth nonlinear models of the whole process.

3.1 *Switching models of human control behaviour*

We frame a finite number of human 'states' as hypotheses. We assume that the human can only be in one 'state' at a time (implying a hard switch between behaviours). Each hypothesis has an associated behaviour which can be described in terms of a probability model. We thus have a standard probabilistic framework for the interpretation of a time series of human action.

To implement this many researchers use the Hidden Markov Model technology from speech recognition research, and much of the framework is taken from the excellent review article by (Rabiner, 1989). This allows us to identify both the parameters of the individual behaviour models, and the switching functions simultaneously. It provides a framework for estimation of a given human 'state' as well as for classification of individual users from their actions. In a related paper, (Yang *et al.*, 1994) applied Hidden Markov Models to learn human skills.

3.1.1. *Model structure* We have a model $\mathcal{M} = \{A_i(x; w_i); f_i(x; \theta_i); \Sigma_i; \pi\}, i = 1..N_m$, with an observable continuous state x and a hidden discrete state q, which can be in one of N_m states. $A_i(x; w_i)$ is the statetransition matrix from discrete state q, dependent on continuous state x, the entries of which $a_{ij} = P(q_{t+1} = j|q_t = 1)$ are the transition probabilities at time t. This is effectively a pattern recognition system, mapping regions of the statespace to a transition probability distribution. In (Murray-Smith, 1998), the transition probabilities are represented by a multinomial logit (or 'softmax') function, with parameters w. The N_m submodels $b(o_t; i)$ define the emission probabilities – i.e. the probability of observing o_t given discrete state $q_t = i$, and continuous state x. In the continuous control action case examined here, one could use a mixture model density function, but in this paper we will only use a single component in each mixture, i.e. a linear model with Gaussian noise (mean $\mu_i = f_i(x; \theta_i) = \theta_i x$, variance Σ_i^2). The estimate of the initial discrete state distribution $\pi = \pi_i$, where $\pi_i = P(q_1 = i)$.

3.1.2. *Inference* We have several inference problems:

(1) *Evaluation:* Probability of model $\mathcal{M}$ given observed output time series $O = \{o_1, o_2, ..., o_T\}$?
(2) *Decoding:* Probability of hidden state i at time t given observed time series O?
(3) *Estimation:* What are the parameters most likely to have generated the output time series?

3.1.3. *Role of inference in modelling human behaviour* The algorithms for inference described above have very concrete uses in a human modelling application. We take the continuous state x to be the input/state information the human bases his or her control on. That control action u is the observable sequence O referred to in the previous section.

- *Evaluation for Classification* There are many applications where we would wish to classify a series of human control actions. The class chosen could be to estimate which of a number of known individual users performed them (possibly for security or insurance purposes), or to compare the behaviour to a number of types of user (e.g. beginner, average user, expert, tired performance). This could be useful in training operators in simulated environments, when classifiers which quantify the style of behaviour could be used to guide and document the results of a training programme. A further example is to differentiate between types of behaviour of a given human operator (e.g. normal behaviour, tired or inattentive behaviour, aggressive behaviour). The approach used is to collect data O_i for each class of control action, and estimate accompanying models $\mathcal{M}_i$. We then select the model with the maximum $P(O|\mathcal{M}_i)$. This is not an explicitly discriminative approach, and if classification is the ultimate aim of modelling, it may be worth using discriminative approaches.
- *Decoding for segmentation of the time-series* The use of standard inference with the model, and EM to iteratively optimise the parameters, automatically gives us a segmentation of the human control timeseries into sub-behaviours. The γ_t's provide us with the probability that the human was in the given state at time t. This is an attempt to infer human 'intentions' or 'sub-goals'. This information can then be used to improve the interaction in a human–machine control system, and can provide context information to human–machine interfaces. Multiple-Model Adaptive Control (MMAC) is an ana-

logue method used in control applications, e.g. (Schott and Bequette, 1997)

- *Estimation for modelling* Given the segmentation of the data provided by the γ_t's, the estimation stage provides us with local models corresponding to system behaviour in each hypothesis. Again, this information, with the γ's can be used to improve cooperation in a human–machine system – we have an estimate of the human's 'hidden state', which can be viewed as current intentions or a current mode of behaviour. We also have how the human usually behaves in this state – the local model associated with that state.

In this approach we examine a model which switches between a number of behaviours. The model switching is probabilistic but conditioned on the state/input vector, so it supports models from purely stochastic switching to purely deterministic switching, depending on its parameterisation. It can also be thought of as a pattern recognition system which chooses the next model state given the 'pattern' of the current continuous state-vector. This probabilistic model can be used to take into account both variations in human behaviour and measurement errors. Note that the multiple-model approaches described in (Murray-Smith and Johansen, 1997) which partition the input space into regions associated with different models are related, but do not contain a stochastic component. The model can be defined from *a priori* knowledge or can be identified from experiments with human subjects.

3.2 *Nonparametric models of human behaviour*

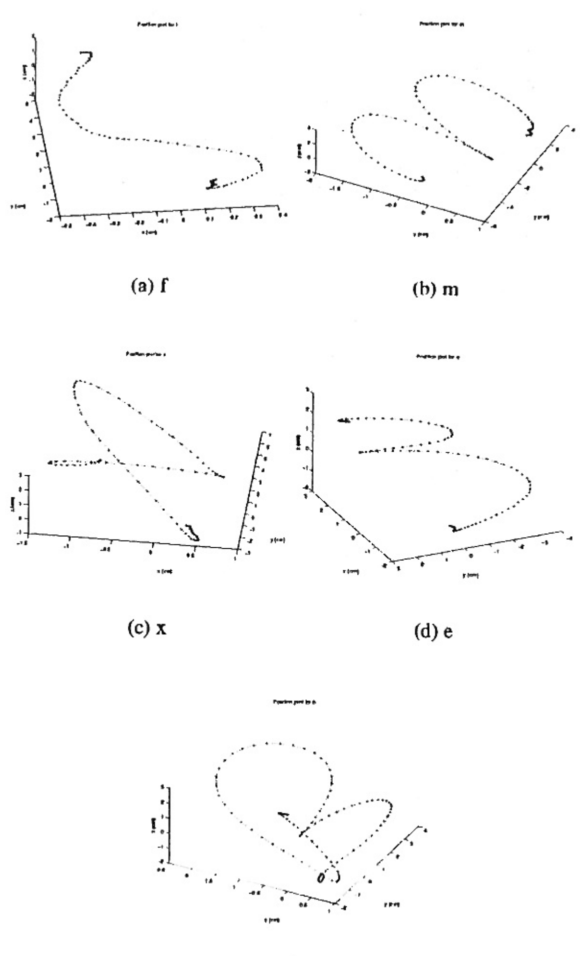
Rather than switching between a number of simple linear sub-models to create the complex behaviour, we could fit a smooth nonlinear model to the data. This could be based on choice of a suitable set of basis functions and regularisation criteria, or could be an inference mechanism such as a Gaussian Process Prior (Williams, 1998; Murray-Smith *et al.*, 1999). Ramsay (2000) would be an example of this approach.

4. ILLUSTRATIVE EXPERIMENTS

The example used in this paper is a specially instrumented hand-held device which provided three-dimensional movement and orientation data. Specific trajectories were defined to be certain symbols or commands for the computer. These had to be recognised automatically by the system without human interpretation.

4.1 *Choice of state and preprocessing*

In this work we measure the x, y, z positions at each time-step t, and the orientation of the object ϕ, θ, ψ (roll, pitch, yaw). Estimation of derivatives in gestures is often extremely useful as the interesting invariants may be far more recognisable in derivatives than in the original states. The first (u, v, w) and second derivatives (a_x, a_y, a_z) of position to be useful, but direct estimation using differences is, however, not a robust strategy, especially in systems with a high sampling rate. Samples are made every 17ms in this example. Figure 5 shows that smooth estimates of derivatives can be obtained with the use of regularisation. It is not immediately obvious that we should use the orientation data, but this, together with the rotational velocities $(\dot{\phi}, \dot{\theta}, \dot{\psi})$ might help disambiguate a number of otherwise similar trajectories.[1] This brings us to a representation similar to that used in, for example, flight dynamics models – this highlights some possible unexpected interactions between fields.



(a) f (b) m (c) x (d) e (e) b

Fig. 3. 3-dimensional plot of 5 different gestures.

Figures 3(a)-3(e) show typical examples of the five gestures used in this test. Figure 4 shows repeated acquisitions of one of the gestures – 'x', and the time-series associated with this gesture are shown in Figure 5.

[1] For space reasons we do not plot the orientation coordinates.

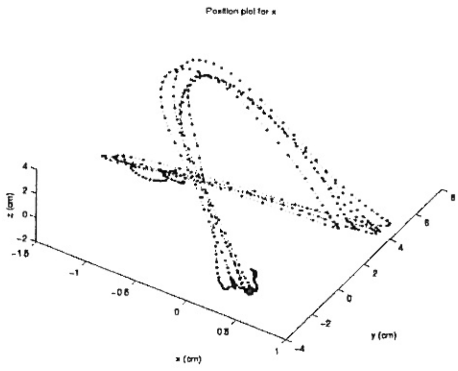
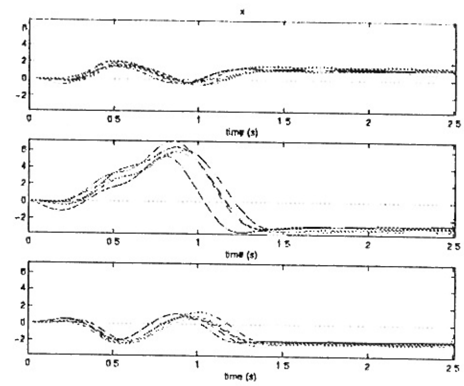


Fig. 4. Example of five 'x' gestures. Dots represent sampled data points, and the dotted lines interpolation between these.

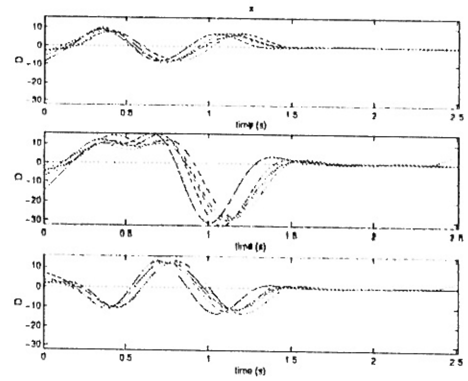
It is important to understand which aspects of the signals measured are important and which vary widely over each gesture. Ramsay and Silverman (1997) describe functional Principal Component Analysis (PCA) approaches which help with visualisation of typical characteristics of variance in collections of time-series.

An important issue is time-invariance. Is a gesture carried out at half the speed still the same gesture? Can we normalise gestures with varying speeds of execution? This implies e.g., a model such as $y(t) = x_i(h_i(t)) + \epsilon_i(t)$, where $h_i(t)$ is the time-warping function. Ramsay and Li (1998) describe this process *curve registration*. Another approach is to assign *landmarks* to points on the gesture and use these to help with alignment, but this is not always straightforward, as markers can be missing from certain curves. Other issues such as rotational or scaling invariance are similar to classical static handwriting analysis.

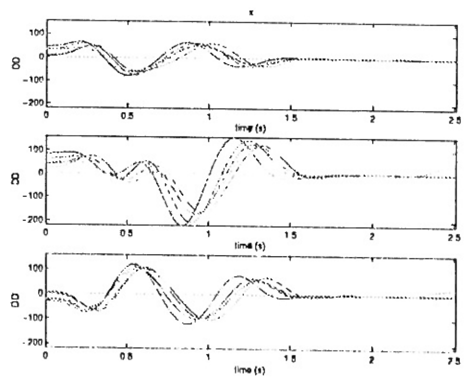
Once a model has been developed we can use it to estimate the probability that a measured time-series corresponds to a given gesture. We can have a straightforward comparison of likelihood, or integrate the approach into a multi-level system, which can bring in contextual information and constraints into the inference. Scores on principal components can be used, as shown in Figure 6. As can be clearly seen, the various gestures fall into distinct clusters, allowing us to distinguish them from each other. The only clear problem case is the 'm' at the top of the graph – on checking this turned out to be a data acquisition problem and looked nothing like the other 'm' gestures. This provides some reassuring insight into the relative robustness of the method to outliers.



(a) Five 'x' gestures



(b) Estimated first derivatives



(c) Estimated second derivatives

Fig. 5. Time-series plot of five 'x' gestures, with estimated derivatives. Without using regularisation, the derivatives would be extremely noisy by the second order.

5. OUTLOOK

This paper presents some exploration of the potential of statistical models in recognising user gestures. Initial results indicate that the use of modern statistical approaches may provide the algorithms needed to begin to make significant progress in this area.

The multiple model framework can be used to identify a model, and estimate the current human 'state' from recorded time-series. This can be used to better coordinate human and machine control behaviour. Murray-Smith (1998) illustrated some of these issues in a simple vehicle control task simulated on a fixed-base PC with joystick and pedals.

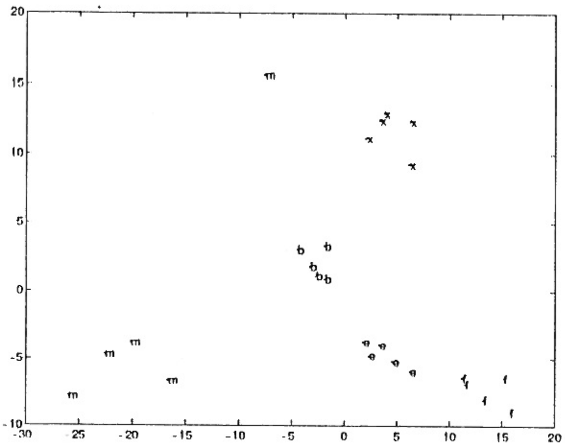


Fig. 6. Example of PCA analysis of gesture data. The figure plots a point for each gesture at the x, y positions corresponding to its score against the two principal components.

In this paper, we develop some of the ideas by basing a gesture recognition system around statistical models. The trajectories created can be related to standard theory for identification of state-space models for flight-dynamics, and this provides a novel insight into the gesture recognition problem, which avoids some of the problems seen in competing non-dynamic approaches.

Functional Data Analysis appears a promising method for data-exploration and analysis in control related human interaction tasks, which are inherently variable.

6. ACKNOWLEDGEMENTS

The author gratefully acknowledges the support of Marie Curie TMR grant FMBICT961369, and the *Multi-Agent Control* Research Training Network supported by EC TMR grant HPRN-CT-1999-00107.[2] Gesture data was provided by Adrian Ng and Ashley Walker of Glasgow University. Some of the plots used the Matlab FDA routines provided by J. Ramsay of McGill University.

7. REFERENCES

Bregler, C. (1999). Learning and recognizing human dynamics in video sequences. In: *Proceedings of IEEE Conf. Comp. Vision and Pattern Recognition, San Jan, Puerto Rico*.

Bregler, C., S. Omohundro, M. Covell, M. Slaney, S. Ahmad, D. A. Forsyth and J. A. Feldman (1998). Probabilistic models of verbal and body gestures. In: *Computer Vision in Man-Machine Interfaces* (R. Cipolla and A. Pentland, Eds.). pp. 318–362. Cambridge University Press. Cambridge.

Harling, P. and A. Edwards (1997). *Progress in Gestural Interaction*. Springer.

Ju, P., L. Pack Kaelbling and Y. Singer (2000). State-based classification of finger gestures from electromyographic signals. In: *Proceedings of International Conference on Machine Learning*.

Murray-Smith, R. (1998). Modelling human control behaviour with a Markov-chain switched bank of control laws. In: *7th IFAC Symposium on Man-Machine Systems, Kyoto*. pp. 521–526.

Murray-Smith, R. and T. A. Johansen (1997). *Multiple Model Approaches to Modelling and Control*. Taylor and Francis, London.

Murray-Smith, R., T. A. Johansen and R. Shorten (1999). On transient dynamics, off-equilibrium behaviour and identification in blended multiple model structures. In: *European Control Conference, Karlsruhe, 1999*. pp. BA–14.

Rabiner, R. L. (1989). A tutorial on Hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE* **77**(2), 257–286.

Ramsay, J. O. (2000). Functional components of variation in handwriting. *Journal of the American Statistical Association* **95**, 9–15.

Ramsay, J. O. and B. W. Silverman (1997). *Functional Data Analysis*. Springer-Verlag.

Ramsay, J. O. and X. Li (1998). Curve registration. *Journal of the Royal Statistical Society, Series B* **60**, 351–363.

Sawada, H., T. Notsu and S. Hashimoto (1998). HMM based gesture recognition using an acceleration sensor. In: *Proceedings of IEEE International Workshop on Robot and Human Communication RO-MAN'98*. pp. 500–506.

Schott, K. D. and B. W. Bequette (1997). Multiple model adaptive control. In: *Multiple Model Approaches to Modelling and Control* (R. Murray-Smith and T. A. Johansen, Eds.). Chap. 11, pp. 269–292. Taylor and Francis, London.

Singer, Y. and N. Tishby (1994). Dynamical encoding of cursive handwriting. *Biological Cybernetics* **71**, 3.

Williams, C. K. I. (1998). Prediction with Gaussian processes: From linear regression to linear prediction and beyond. In: *Learning and Inference in Graphical Models* (M. I. Jordan, Ed.). Kluwer.

Wilson, A. and A. Bobick (1995). Using configuration states for the representation and recognition of gesture. Technical Report 308. M.I.T. Media Laboratory Perceptual Computing Section.

Yang, Jie, Yangsheng Xu and Chiou S. Chen (1994). Hidden Markov Model approach to skill learning and its application to telerobotics. *IEEE Transactions on Robotics and Automation*.

[2] This work is the sole responsibility of the author, and does not reflect the European Community's opinion. The Community is not responsible for any use that might be made of data appearing in this publication.