How variance in parameters affect the reproduction number and why the SEIR model overshoots

The modelling of COVID19 has become a critical aspects of present life. In this fundamental task, a standard procedure has become the implementation of more or less detailed differential equations, that estimates the time development in the number of susceptible, exposed, infected, and recovered (SEIR) individuals. However, these SEIR models are based on very simple assumptions, which constitute obvious approximations. In this paper, we introduce an agent based network (ABN) model, that allows us to test the impact of two fundamental assumptions of the SEIR models: spatial homogeneity and infection rate homogeneity. Based on Danish data, we attempt to put realistic estimates of their impact on both the initial predictions and also of the long term development. Our results suggests (?) that while spatial homogenity is of minor importance, the rate homogenity assumption has a major impact on the long term development, suggesting a factor 2 (?) difference in the sum of infected individuals before the epidemic ends.

Keywords: Stochastic Networks, Epidemics, Numerical predictions.

Introduction

The dynamic of a single epidemic of a disease is often described in terms of the susceptible-exposed-infected-removed (SEIR) model and its related variants. The main assumption of this model is that the population, in which a pathogenic agent is active, comprises three subgroups, described by the differential equations: $dS/dt = -\beta SI$, $dI/dt = \beta SI - \gamma I$ and $dR/dt = \gamma I$. In this model, individual members of the population susceptible to disease (S) become infected (I), and are subsequently removed (R) from the pool spreading the disease due to either their death or newly acquired immunity. This SIR model has been widely used to describe infectious diseases and presently being a core part of every countrys response to the COVID 19 outbreak. The model examines only the temporal dynamics of the infection cycle and can be improved by including an exposure (E) state and various other modifications, which leads to more realistic (mSEIR) models. In mSEIR models, some fundamental approximations has been made that are in some systems valid. These are that the system is well-mixed (spatial homogenity), meaning that all rates of infection between nodes are the same, and that every node on average has the same number of connections (infection rate homogeneity). However, it is more than doubtful, that these assumptions are fulfilled in the current pandemic, and they remain a source of bias, the size of which can not be determined within the mSEIR regime. In the following, we compare an mSEIR model to a corresponding agent based network (ABN) model to quantify the differences in early behaviour (from which predictions are made) and long term outcome.

Models considered

Modified Susceptible-Exposed-Infected-Recovered model

We have chosen to consider an mSIER model (see Figure 1???), which consists of one Susceptible state (S), four Exposure states (E_{1-4}), four Infection states (I_{1-4}), and one Removed (R) state. The rate of transition through these

states is governed by the following formulae (schematised in Figure 1a???):

$$\begin{aligned} \mathrm{S} &\xrightarrow{\beta_{ij}} \mathrm{E}_1 \\ \mathrm{E_k} &\xrightarrow{\lambda_E} \mathrm{E_{k+1}} \\ \mathrm{E}_4 &\xrightarrow{\lambda_E} \mathrm{I}_1 \\ \mathrm{I_k} &\xrightarrow{\lambda_I} \mathrm{I_{k+1}} \\ \mathrm{I}_4 &\xrightarrow{\lambda_I} \mathrm{R} \end{aligned}$$

The dynamics of the infections are described with the differential equations shown below:

$$\dot{S} = -\beta_D SI \quad \text{with } \beta_D = \beta\frac{\mu}{2N_0} \tag{1}$$

$$\dot{E}_1 = \beta_D SI - \lambda_E E_1 \tag{2}$$

$$\dot{E}_i = \lambda_E E_{i-1} - \lambda_E E_i \quad (i = 2, 3, 4) \tag{3}$$

$$\dot{I}_1 = \lambda_E E_4 - \lambda_I I_1 \tag{4}$$

$$\dot{I}_i = \lambda_I I_{i-1} - \lambda_I I_i \quad (i = 2, 3, 4) \tag{5}$$

$$\dot{R} = \lambda_I I_4 \tag{6}$$

This constitutes a deterministic model, with a combination of classical elements. The choice of four exposure and infection states is motivated by the observed transition dynamics of the Covid-19 virus [Ref: ?], but not an essential part. The purpose of the mSEIR model is as stated to be able to compare it with an agent based network model.

Agent based network model

We consider an agent based network (ABN) model of N_0 nodes, each representing an individual in a population. We will assume that each node, i, has a number of connections n_i, and that each link between node i and j has an assigned interaction strength $\beta_{ij} = \beta_{ji}$. For this network, the average degree of connectivity is μ, and the network is constructed by generating $\mu \cdot N_0$ links.

From this we simulate an epidemic, starting from N_{init} initial points. In the ABN model, we assume that each node will make the *same transitions* between exposed and infected states, as in the mSEIR model, Eq. . In the corresponding spread on the network, an infected node can only infect the subpopulation of nodes, which it interacts with. We will now compare this to the structure of a typical SEIR model. The SEIR model is deterministic and assumes a well mixed population. The corresponding model for a population of N_0 individuals take the form:

$$\dot{S} = -\beta_D SI \quad \text{with } \beta_D = \beta\frac{\mu}{2N_0} \tag{7}$$

$$\dot{E}_1 = \beta_D SI - \lambda_E E_1 \tag{8}$$

$$\dot{E}_i = \lambda_E E_{i-1} - \lambda_E E_i \quad (i = 2, 3, 4) \tag{9}$$

$$\dot{I}_1 = \lambda_E E_4 - \lambda_I I_1 \tag{10}$$

$$\dot{I}_i = \lambda_I I_{i-1} - \lambda_I I_i \quad (i = 2, 3, 4) \tag{11}$$

$$\dot{R} = \lambda_I I_4 \tag{12}$$

This is schematically shown in Figure 1A. Our goal is to test the agreement between the deterministic mSEIR model and a stocastic ABN model. Especially we wanted to estimate the effect from spatial inhomogeneties and infection rate inhomogeneties. Before we simulate this system, we generate a fundamental network of connections (Figure 1B) where we use the general algorithm:

Result: Connect Nodes

initialization;

for i = *1:N* **do**

  W(i) = $\frac{-log(rand)}{\mu_W}$;

  Age(i) = int(10*rand) ;

end

$PP = np.cumsum(W)/np.sum(W)$;

Connect nodes;

for i = *1:N*μ_N* **do**

  ra1 = rand;

  id1 = np.searchsorted(PP,ra1);

  accept = 0;

  while *accept ==0* **do**

    ra2 = rand;

    id2 = np.searchsorted(PP,ra2);

    rAge = rand;

    rDist = rand;

    $r_{ij} = \sqrt{(px[id1] - px[id2])^2 + (py[id1] - py[id2])^2}$;

    if $AgeMat[Age[id1], Age[id2]] \leq rAge$ *and* $rDist \leq e^{-\rho r_{ij}}$ **then**

      connect nodes!

end

end

end

Algorithm 1: Generation of the network

To simulate the disease on the network, we use an event-driven Gillespie algorithm. To optimize speed and precision every time a new node is infected, this element is removed from all his connections, so we avoid the same person being infected (or attempted to be infected) twice. At each event we decide whether a node should move up one state or an infected node should pass on the infection to one of his neighbours. All this is selected relative to all the rates of every event.

Comparison of mSEIR and homogeneous ABN models

Before incorporating more detailed and realistic assumptions into the ABN model to see their impact, we first want to ensure that the ABN model can reproduces the mSEIR model when based on the same assumptions of spatial and infection rate homogenity. To do this, we use the connection algorithm but set $W = 1, AgeMat = \mathbf{1}$ and $\rho = 0$. To minimise the effect of statistical fluctuations, we set $N_{init} = 1000$, when comparing the mSEIR and ABN models. To study the effects of the ABN model, we measure the difference in the height of the curves of number of infected as

shown in Figure 1C. We note that the mSEIR model rises a little earlier (see Figure 1C), which can be understood in terms of the continuous nature of the model, where the ABN has integer requirements. We also tested this as a function of μ and found that this result is decaying as well (Figure 1ED. We find that the two resemble each other well if μ is large, but at very small values of μ there is a fundamental difference and the ABN shows a significantly lower result. We tested this as a function of μ and found that this effect becomes negligible when $\mu > 10$ (Figure 1E). Next we investigated the effect of the size of the network in the ABN model. By increasing the number of nodes in general, the difference between the mSEIR and ABN is decreasing, and the time difference is decreasing (Figure 1F), but the height of the peak and the overall integral (R_∞) in the ABN model is smaller by 3% (?) and 4%, respectively (Figure 1G). Finally we tested the value of beta, i.e. the interaction strength between two connected nodes. Here we found that this does not significantly affect the results (Figure 1H+I) The precision of the ABN model and the impact of the ramdomness can be investigated by running the ABN model 1000 times. Figure 1C shows the result of this. As an important question we wanted to investigate the following: In diseases, we use models to predict the outcome, given an initial set of datapoints. Given we know the number of infected (i.e. everybody is tested) what is the expected error in our predictions. To do this, we fitted the data each day between more than 100 people where infected to the point where 1/1000 in the population had been infected.

Having assured ourselves, that the ABN can indeed reproduce the SEIR model with high precision, our next goal is to include some effects that are present in the spreading of a disease, but which are not included in the SEIR model, and determine their impact.

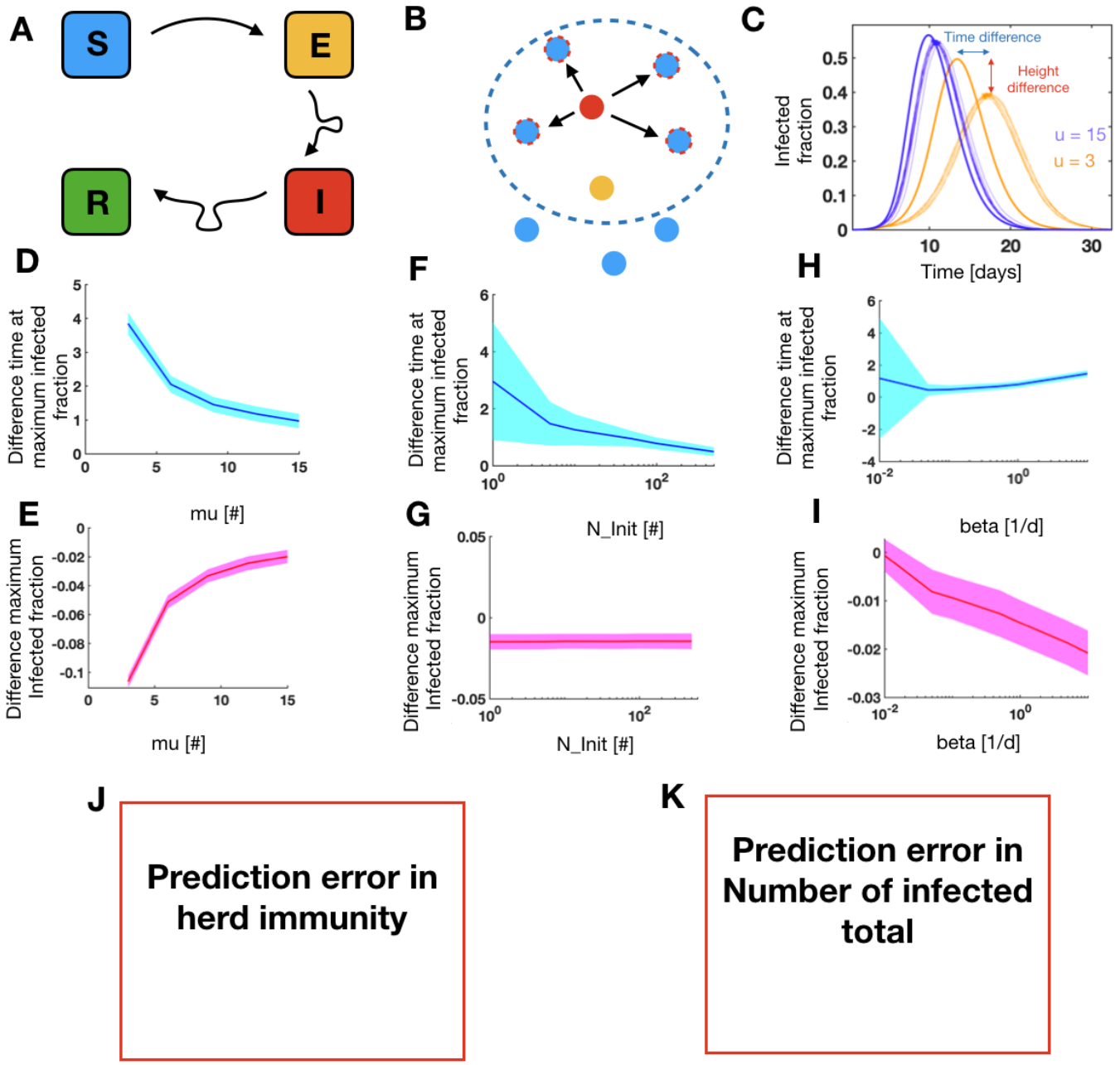
Inhomogenities in node connectivity and interaction strength

Next we wanted to investigate the effect of inhomogeneities in the number of conections and the interaction strength. Each node is therefore assigned a connectivity weight, either fixed (as assumed in the mSEIR model) or following a (shifted) exponential distribution $p(N) = \exp(-(N - \mu/10)/\mu)/\mu, N \geq 4$, as indicated by the DTU-MobilePhone data [Ref:]. For connecting the nodes, pairs of nodes i, j are randomly selected proportional to their connectivity weight. The selection of node pairs is repeated until the mean number of connections reaches $\mu = 40$. The resulting distributions of connections for the nodes can be see in Figure 2G. Doing this we find that the spread in number of connections leads to a huge increase in the number of infected in the SEIR model. The results are shown in Figure 2A-F.

Introducing the effects of space and applying this to a realistic and clustered data set

In order to include more realistic assumptions about the spatial distribution of the population (i.e. nodes), we distribute the network nodes according to housing sales in Denmark 2007-2019 [Ref: Boligsiden]. The distribution has been approximated with a 2D kernel density estimate, from which we randomly select node locations. This very inhomogenious distribution represents a realistic mix of urban and rural population (at least for Denmark).

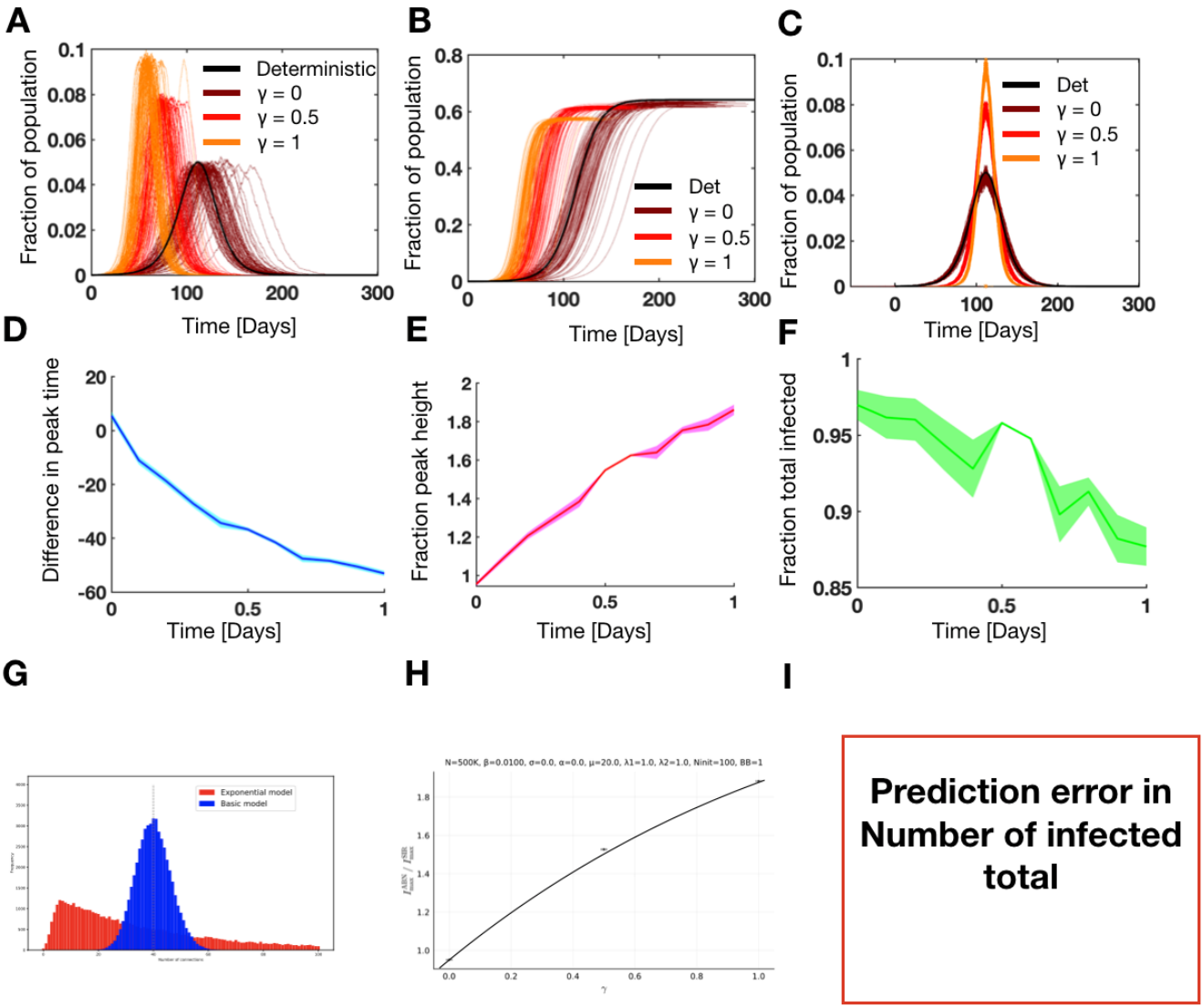
They are then assigned a connection with probability $p(d) = \exp(-\rho r_{ij})$, where r_{ij} is the node distance and ρ is a parameter that reflects the typical connection distance. This reflects the deminishing probability of connecting with distance. The exponential assumetion is not central, as we simply want to see the effect of a change away from the naive mSEIR model. However, in a small fraction of the cases (denoted ϵ_ρ), the two nodes are connected, to include



the effect of long distance connections. This metric means that one is more inclined to infect those in the vicinity than those at a distance, but it does not rule out spread at longer distances. This algorithm leads to urban population being much more interconnected than rural population. If one chooses to keep the first randomly selected node until connected, then the distribution of number of connections become more alike for urban and rural populations. For every attempts, ρ is decreased by 0.05% to ensure an eventual connection.

If one assumes that all people infect at the same rate (as the SIR model does), then the resulting distribution of number of actual infections per person in the ABN model becomes Poisson distributed. However, it is unlikely that this is the case, and there are eeping the *average number of infections per person constant*

All the parameters introduced in the following are listed in table **??** along with those common to the two models. Spatial inhomogenities introduces a bias to interact with and potentially infect your nearest neighbours. When



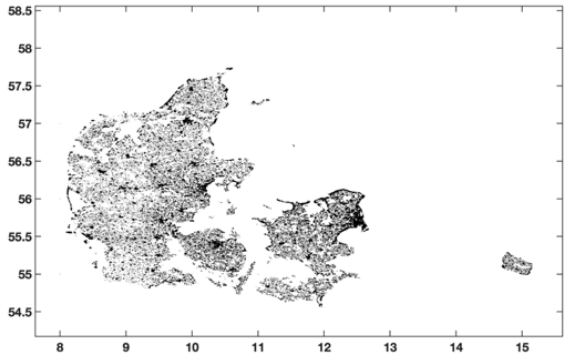
including this effect, we observe that the strength of the epidemic diminishes by ? %.

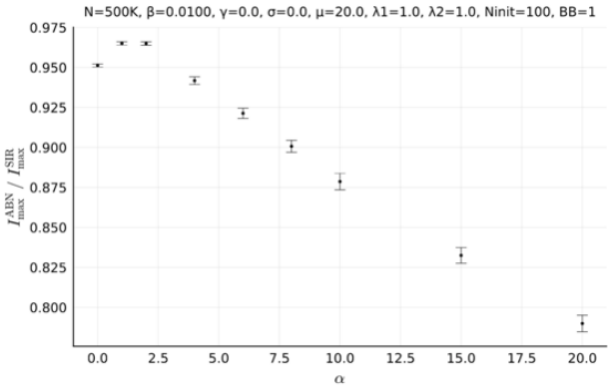
Combination of effects and how to account for this in the modelling

Having considered how each of the two inhomogenity assumptions each affect the...

Hospitalisation of infected

To model Hospitalization we include 3 new states, H (in hospital), ICU (in intensive care) and D (dead). When a node reaches the last step (I_4), then it is placed in the recovered state with an age dependent probability, otherwise it is placed in the state H. We then introduce a matrix, giving the probabilities to move between states:





Prediction error in herd immunity

Prediction error in Number of infected total

Parameter	Concept and Description
N_{tot}	Total population size in model
N_{init}	Initial number of infected individuals
β	Rate of Infections $S \to E_1$
λ_E	Rate of transitions out of E states
λ_I	Rate of transitions out of I states
μ	Average number of connections
σ_μ	Fixed or exponentially distributed number of connections (formerly σ)
ρ	Exponent of exponential infection rate with distance (formerly α)
σ_ρ	Constant or exponentially distributed infection rates with distance (formerly σ)
ConnectAlgo	Algorithm for connecting nodes (with and without replacing the first node during connection attempts)

TABLE I: List of parameters used for the models. The first six parameters (above the middle line) are common to the mSEIR and ABN models, while the last four (below the middle line) are specific to the ABN model, and represents the introduction of network and inhomogeneties.

$$\text{HosMat} = \begin{array}{c} \\ H \\ ICU \\ R \\ D \end{array} \begin{array}{c} \begin{array}{cccc} H & ICU & R & D \end{array} \\ \begin{pmatrix} 0 & p_{12} & p_{13} & p_{14} \\ p_{21} & 0 & p_{23} & p_{24} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

These probabilities are all age dependent, and extracted from the Norwegian health institute (https://www.fhi.no/en/id/infectious-diseases/coronavirus/coronavirus-modelling-at-the-niph-fhi/)

Conclusions

I. SUPPLEMENTARY MATERIAL