# Aggregated Markov Models of a Heterogeneous Population of Photovoltaic Panels

A. Peruffo<sup>1</sup>, E. Guiu<sup>2</sup>, P. Panciatici<sup>2</sup>, and A. Abate<sup>1</sup>

<sup>1</sup> Department of Computer Science, University of Oxford, UK <sup>2</sup> Réseau de Transport d'Electricité, Paris, France

**Abstract.** We present a new framework for aggregated quantitative modelling of a heterogeneous population of photovoltaic panels. We are interested in the behaviour of photovoltaic panels as electric power sources, and in an aggregated model that can capture how such a population behaves when connected to the power grid. After an initial analysis of the characteristics and behaviour of a single device, we propose two Markov chain models for the aggregation of a heterogeneous population of such devices. We study the dynamical behaviours of the aggregated models, embedded within the dynamics of the grid frequency. A simulation study shows the effectiveness of the aggregated models when compared to the physical system, and leads to conclude that population heterogeneity is a desirable feature for the overall system dynamics.

### 1 Introduction

In recent years, both academia and industry have increased their interest and attention on renewable energy sources. The growing trend toward environmental preservation, witnessed by the famous Kyoto protocol, the Paris Agreement and the so called 2-degrees challenge, is leading scientists to focus on new technologies and their applications in power generation. Well known technologies are e.g. wind power, solar energy and geothermal energy. Devices using these energy sources are typically distributed over a large area rather than being concentrated in a small production area, which leads to consider the issue of distributed generation. In the field of renewables, many studies are now focused on photovoltaics (PV), which is nowadays the third most important renewable energy source, after hydro and wind power, in terms on total installed capacity [1]. PV panels produce electrical current from solar irradiation by virtue of the photovoltaic effect of semiconductor materials. In this work we will interchangeably denote photovoltaic panels as PV or solar panels.

Although models for a single PV cell or panel are well known in literature [2], up to the authors' knowledge there has not been any model encompassing the connection between panels and grid. Large PV farms have been modelled and studied as a whole, but a model dealing with distributed generation of power in a large city, or a country, from household applications is still missing. This can be due to many reasons: first of all because photovoltaic application for a single home has a limited power to inject into the grid and therefore can be neglected; secondly because often the power produced is consumed at the source (the house itself), hence the net contribution to the wider grid is nearly null.

Nevertheless, the growing population of PV panels justifies the study, modelling and control of this energy source. The power network, at a regional or state level, must be prepared to deal with the volatility and unpredictability intrinsically related to the production of renewable energy. As an example, PV generated power – distributed over 1.5 million PV setups – has provided for approximately 7.5% of Germany's electricity demand, with peaks of 50% during weekends and holidays [3]. Often the power production of such devices do not follow usual demand patterns: the power production of a panel, in a clear day, follows the irradiance of the sun, with a maximum at around midday, when the power is usually injected into the grid, due to a lack of consumption. This unbalanced flow, if not handled correctly, could lead to network issues, such as blackouts [4]. Great care has been taken to proactively cope with the eclipse occurred in March 2015: it was anticipated that a reduction of around 30 *GW* could have been caused by the sun occlusion [5], which is about 10 times the size of a blackout accident that can be resiliently handled by the grid.

In this paper, we present a new framework for modelling and abstraction of a large population of photovoltaic panels. We firstly analyse the behaviour of a single physical device when connected to the grid. The PV panel is equipped with a sensor, to sample the network frequency, and with an internal counter in order to ensure that the network frequency remains inside a certain range of admissible values for a defined amount of time, before an action is taken. Two quantities seem to be key to model the PV behaviour: the working interval of network frequencies, and the internal time delay required for a safe connection to the network. Each device in principle can have a different admissible frequency range and a different time delay. In order to model this heterogeneity we present two discrete-time Markov chains, one closer to the physical description with (n+2) states, and the other more abstract with 3 states. These models are then connected to the dynamics of the network frequency, giving an extended model for the whole system, for which stability properties are studied.

The remainder of the paper is organised as follows. In Section 2 we describe the behaviour of a single photovoltaic panel in frequency. The Markov chains models and the frequency evolution are explained in Section 3. The performance comparisons between the two models and a realistic model taken as benchmark in Section 4. Conclusions are drawn in Section 5.

# 2 Description of the physical system

In order to provide an aggregated model for a population of PV panels, we start with the description of the behaviour of a single panel, as a function of the local dynamics of the grid.

A panel has two working states, ON and OFF. It switches amongst these two states according to two conditions: the local network frequency  $f(\cdot)$  (whose nominal value is taken as  $f_0 = 50 \ Hz$ ) and an internal time delay  $\tau_r$ , usually given in seconds (cf. Table 1). The panel is connected to the grid and senses it by sampling its frequency discretely in time. It can be in the ON mode if the network frequency lies within a given local frequency interval  $\mathcal{I}_f$ , otherwise it must disconnect and transition onto the OFF mode. It is sensible to assume that the ON-to-OFF switch happens within a negligible time interval [6, 7], whereas the OFF-to-ON switch cannot happen before a time delay  $\tau_r$ , during which the frequency f must dwell within  $\mathcal{I}_f$ : this is in order to ensure that the network frequency is stable enough to render the panel connection to the grid safe, and to avoid chattering behaviours that can lead to overall network instability. During this interval of time the panel keeps sampling the frequency and if it measures it to be outside the working interval, the internal counter is reset. In order to describe this functioning, one can imagine a PV panel as a device equipped with an internal counter  $\tau(k)$  at time k and a time threshold  $\tau_r$ . This counter is set off when the device is in the OFF state and the frequency enters the working interval  $\mathcal{I}_f$ . If the counter value reaches  $\tau_r$  as  $f \in \mathcal{I}_f$ , then the device can turn to the ON state. If instead the frequency goes outside  $\mathcal{I}_f$  while  $\tau(k) < \tau_r$ , then the device resets its counter. As the PV panel senses the network frequency via a digital sensor with a defined sampling rate,  $\tau_r$  is a value given in number of samples. In practice, the sampling time of a PV panel is in the order of 200 ms, and  $\tau_r$  is around 20 s. The internal clock  $\tau$  can then be thought of as a counter, and as such it will be modelled in this work.

Current state $q(k)$	Frequency	Delay	Next state $q(k+1)$
OFF	$f(k) \in \mathcal{I}_f$	$\tau(k) \ge \tau_r$	ON
ON	$f(k) \in \mathcal{I}_f$	n.a.	ON
ON	$f(k) \notin \mathcal{I}_f$	n.a.	OFF
OFF	$f(k) \in \mathcal{I}_f$	$\tau(k) < \tau_r$	OFF

Table 1: Behaviour of single photovoltaic panel within the power network at time k. Key quantities are: panel state  $q \in \{ON, OFF\}$ ; network frequency f; operating frequency band  $\mathcal{I}$ ; clock/counter  $\tau$  and re-connection delay  $\tau_r$ .

The power injected into the grid by the population of panels has non negligible dynamical effects to the network frequency, which directly influences the behaviour of single panels in a feedback loop.

Our aim is to develop a model for a large population of photovoltaic panels. This cannot be modelled considering N identical PV panels, simply because in reality it is not the case: different regulations, manufacturers, makes and age, all render the population highly heterogeneous [6–8]. As such, this work focuses on parameters heterogeneity as a key aspect: we deal with network frequency thresholds and time delays, to model the more realistic situation where each panel has a different working interval as well as a different counter. Another interesting part of the real system is its power output: the production of electrical power is subject to many external factors, e.g. weather conditions, light occlusions. However, here we will focus on a power production that is constant

in time. Time varying features can be added at a second stage, for example modelling a power production as a stochastic process [9].

# 3 Markov chain model of a population of PV panels

Since the behaviour of each panel depends on a discrete sampling of the frequency, we refer to a discrete-time framework and model the aggregation of photovoltaic panels with a discrete-time Markov chain. Since the focus is on the frequency thresholds and the delays of the panels, for simplicity we consider the power production to be at its maximum nominal value whenever a panel is ON.

As an early simplifying assumption (to be shortly repealed in the next subsection), assume exact population homogeneity, meaning that every device shows the same behaviour in time and is characterised by the same parameters. This allows us to define a quantity P expressing the weighted power production as

$$P = \frac{1}{N} \sum_{i=1}^{N} P_i,$$

where N is the total number of panels in the population, and  $P_i$  is the power output of the single *i*-th panel. We can consider the normalised power production R(k), at time k, as

$$R(k) = \frac{1}{NP} \sum_{i=1}^{N} q_i(k) P_i,$$

where  $q_i(k) \in \{0, 1\}$  denotes whether the *i*-th device at time k is in the OFF or ON state, respectively. Since we have assumed population homogeneity, all the devices behave in accord, and  $R(k) \in \{0, 1\}$  at each time step k: we can then consider R(k) as a Bernoulli random variable, and introduce

$$x(k) = E[R(k)] = \mathbb{P}[R(k) = 1],$$

a variable defined as the expected value of R(k) at time k, which denotes the probability of being in the ON state at that time. Furthermore, by the law of total probability,

$$\begin{aligned} x(k+1) &= \mathbb{P}[R(k+1) = 1] = \mathbb{P}[R(k+1) = 1 | R(k) = 1] \cdot \mathbb{P}[R(k) = 1] + \\ &+ \mathbb{P}[R(k+1) = 1 | R(k) = 0] \cdot \mathbb{P}[R(k) = 0]. \end{aligned}$$

Let us define  $a(k) = \mathbb{P}[R(k+1) = 0|R(k) = 1]$ , so that  $\mathbb{P}[R(k+1) = 1|R(k) = 1] = (1 - a(k))$ , and let us introduce  $b(k) = \mathbb{P}[R(k+1) = 1|R(k) = 0]$ : the previous equation can be rewritten as

$$x(k+1) = \mathbb{P}[R(k+1) = 1] = (1 - a(k))x(k) + b(k)(1 - x(k)).$$
(1)

This relation describes how the probability of being ON gets updated at time k. In the framework that we have adopted, the transition probability ON-to-OFF (a(k)) and OFF-to-ON (b(k)) are governed by the value of the network

frequency f(k), namely whether or not  $f(k) \in \mathcal{I}_f$ . Since we assumed population homogeneity, these values are binary: for instance, when  $f(k) \in \mathcal{I}_f$  then a(k) = 0whilst b(k) = 1. Since the power output is the sum of the PV devices turned ON, it can be expressed as  $P_{out}(k) = NPx(k)$ .

### A Markov model without delays

Let us now introduce heterogeneity over the frequency behaviours of different panels, as expected in reality. Let us suppose that each panel has different frequency thresholds (which we take to be constant in time): each panel reacts to the network frequency distinctively, namely it disconnects/reconnects at different frequencies than other panels. We assume that these thresholds are distributed continuously according to a known probability distribution - such continuous statistics can be regressed from discrete population data. Similarly to the homogeneous case, we introduce b(k) and a(k) as the probability of turning to the ON state or to the OFF state at the (k + 1)-st time step, starting from the opposite condition at the k-th time step, respectively. Unlike the homogeneous case where a and b were binary, in order to encompass population heterogeneity we integrate the probability distribution function comprising the frequency thresholds using the current value of network frequency f(k) as one of the extrema: we thus obtain the portion of panels that are enabled to change their state. Formally,

$$a(k) = \int_{-\infty}^{f(k)} p df^{0|1}(u) du, \qquad b(k) = \int_{f(k)}^{\infty} p df^{1|0}(u) du,$$

where  $pdf^{0|1}$  is the probability density function of a random variable modelling the transition from R(k) = 1 to R(k + 1) = 0, and analogously for  $pdf^{1|0}$ . Alternatively, a(k) and b(k) can be expressed via the cumulative distribution function of the known probability distribution for the frequency thresholds.

The expression in (1) can be interpreted as a Markov chain with two states and time varying transition probabilities, as depicted in Figure 1. Here the edges representing the transitions between state ON and OFF are labelled with a(k)and b(k), while the self loops are simply 1 - a(k) and 1 - b(k).



Fig. 1: A time inhomogeneous Markov chain model for the aggregated dynamics, without delays.

#### A Markov model with delays

We now introduce a framework to encompass delays in the aggregate model (cf. Figure 2): as observed in practice, panels cannot switch to the ON state instantaneously. As stated above, each panel has an internal counter for the OFF-to-ON transition. We assume to sample the delays with a coarse sampling time of 1 s: in this way we can lump together the panels which show delays in the same sampling time. We then obtain probability transition values  $\tau_i$ , which represent the probability of switching on after i seconds. In order to pin down this idea, we utilise n states, defined as  $w_i$   $i = 1, \ldots, n$ , representing the *i*-th time step when the network frequency sampled by the panel is within the given threshold, but when the panel has still to turn to the ON state. In other words, the  $w_i$  state describes a device that has been waiting to turn on for *i* time steps. In view of the discussions in Section 2, we focus on the case  $n \gg 1$ . In the *i*-th delay state, there are three outgoing transitions: one towards the ON state, a second towards state i + 1, and one back to the OFF state. The probability associated with the third transition is 1 - b(k), which is the probability of sensing the network frequency outside the working interval. The first outgoing probability is  $\tau_i b(k)$ :  $\tau_i$  is the probability to have a time delay that permits the panel to go from state  $w_i$  to state ON, which can happen only if the frequency is within the working interval (hence the multiplication by b(k)). We have tacitly assumed that the probability distributions of the frequency thresholds and time delay are independent. There can be, in reality, correlation between these two quantities, in which case we need to compute integrals of joint probability distributions.

We assume that  $\forall i, \tau_i \geq \tau_{i+1}$  and that  $\sum_i \tau_i = 1$ , so the terms resemble a geometric distribution that can describe an arrival process or a waiting-time random variable.



Fig. 2: A Markovian model for the aggregated dynamics, with delays.

The dynamics of the Markov chain in Fig. 2 can be summarised as

$$\begin{cases} x(k+1) = (1-a(k))x(k) + b(k)\sum_{i=1}^{n}\tau_{i}w_{i}(k) \\ w_{1}(k+1) = b(k)\left[1-x(k)-\sum_{i=1}^{n}w_{i}(k)\right] \\ w_{i}(k+1) = b(k)(1-\tau_{i-1})w_{i-1}(k) \\ w_{n}(k+1) = b(k)\left[(1-\tau_{n-1})w_{n-1}(k) + (1-\tau_{n})w_{n}(k)\right], \end{cases}$$
(2)

where x(k) represents the probability of being in the ON state at time k. We use this value as the portion of panels ON at time k; similarly,  $w_i(k)$  is the portion of panels waiting to turn ON for i time steps at time k; and a(k), b(k)are the cdf's of the distributions of frequency thresholds in the population of panels. Let us further explain the details of the model in (2). To model real-life applications, we deal with an interval of time in which no panel switches on and a subsequent time interval in which panels switch on according to a geometric probability distribution. The latter time interval is described above with the  $w_i$ states. In order to encompass the time interval in which panels do not switch on (which is around 20 s in real-life applications) we must include new states that are "pure waiting states", denoted  $pw_i(\cdot)$ . Their number depends on the desired delay and on the given sampling time: e.g. if the sampling time is 1 s and the minimum desired delay is 20 s we add 20 new states. These states are such that  $\tau_i = 0$ , so to prevent the possibility to switch on, which boils down to a time delay, as desired. These new equations do not invalidate the previous analysis and for simplicity we will continue our analysis without them.

#### Abstraction of the Markov model with delays

Towards a simplified and more insightful analysis of the dynamics of the model with delays, we aggregate the *n* waiting states into a single location, which thus represents the sum of the portion of devices that are waiting to turn ON and is associated with a new variable  $y(k) = \sum_i w_i(k)$ . To express the overall dynamics, we rewrite the term  $\sum_i \tau_i w_i(k)$  as a function of y(k), and introduce a term  $\varepsilon_k \in [0, 1] \forall k$ , such that

$$\sum_{i=1}^{n} \tau_i w_i(k) = \varepsilon_k \sum_{i=1}^{n} w_i(k), \quad \text{thus} \quad \varepsilon_k = \frac{\sum_{i=1}^{n} \tau_i w_i(k)}{\sum_{i=1}^{n} w_i(k)}.$$

The model now presents only three states, as depicted in Figure 3, whose transition equations are

$$\begin{cases} x(k+1) = (1-a(k))x(k) + b(k)\varepsilon_k y(k) \\ y(k+1) = b(k)(1-x(k)) - b(k)\varepsilon_k y(k) \end{cases}$$
(3)

The new model is smaller and easier to analyse. However, in general we do not know the exact value of  $\varepsilon_k$ , so we seek a value for it that ensures that the error between the two models decreases to zero with time. Define the abstraction error e(k) as the difference between the element x(k) of each model: e(k) =



Fig. 3: Abstraction of the Markov model with delays for aggregated dynamics.

 $x^{n-s}(k) - x^{3-s}(k)$ , where  $x^{n-s}$  and  $x^{3-s}$  denote the x component of the model with n delay states and that with 3 states, respectively. We obtain

$$\begin{aligned} |e(k+1)| &= |x^{n-s}(k+1) - x^{3-s}(k+1)| \\ &= |(1-a(k))x(k)^{n-s} - b(k)\sum_{i=1}^{n}\tau_{i}w_{i}(k) - (1-a(k))x(k)^{3-s} + b(k)\varepsilon_{k}y(k)| \\ &= |(1-a(k))e(k) + b(k)d(k)| \le (1-a(k))|e(k)| + b(k)|d(k)|, \end{aligned}$$

where  $d(k) = \varepsilon(k)y(k) - \sum_{i=1}^{n} \tau_i w_i(k)$  Notice that  $(1 - a(k)) \in [0, 1] \forall k$ , by definition, which results in a term that does not increase with time. We study the evolution of the second term, d(k): whilst  $d(k) \in [-1, 1]$  by definition, we do not know its sign, which could change at each time step k. It can be shown that |d(k)| is a contracting map if we select the value  $\varepsilon_k = \tau_1$ . This result can be interpreted recalling the meaning of  $\varepsilon_k$ : it represents a weighted convex combination of the  $w_i$ ; considering  $\varepsilon = \tau_1$  results in the worst-case scenario, where we utilise the maximum value among the  $\tau_i$ .

In other words, choosing  $\varepsilon_k$  to be the constant value  $\tau_1$  asymptotically decreases the error e(k) between the two models to zero. Note that, knowing the values  $\tau_i$ , we could estimate  $\varepsilon_k$  by estimating the states  $w_i(k)$ . The estimation of the states  $w_i(k)$  can be attained, as will be discussed in Appendix A.

We now illustrate more clearly how the frequency value affects the Markov chains: a(k) can be written formally as

$$1 - a(k) = \int_{f(k)}^{\infty} p df^{0|1}(u) du = 1 - c df^{0|1}(f(k)),$$

and similarly for the term b(k). In order to simplify this nonlinear term, let us linearise it as

$$1 - a(k) = 1 - cdf^{0|1}(f(k)) \simeq 1 - (k_1 f(k) - k'_2) = -k_1 f(k) + k_2,$$
  
$$b(k) = 1 - cdf^{1|0}(f(k)) \simeq -k_3 f(k) + k_4,$$

where  $k_2 = k'_2 + 1$ . We then obtain terms that are easier to study, which allows us to derive conditions to guarantee the stability of the whole system. Note that this approximation will require saturation within the interval [0, 1], and that this is the *exact* form of the cdf in case of a uniform distribution. Model (3) then becomes

$$\begin{cases} x(k+1) = (-k_1 f(k) + k_2) x(k) + (-k_3 f(k) + k_4) \varepsilon_k y(k) \\ y(k+1) = (-k_3 f(k) + k_4) (1 - x(k) - \varepsilon_k y(k)) \end{cases}$$
(4)

### Overall closed-loop model

As previously mentioned, the overall dynamical system under study comprises both the population of panels and the network frequency. So far, we have built aggregated models only for the population of solar panels: we are interested in studying how the network frequency is influenced by the power production of PV panels. We focus on an approximate version of frequency dynamics, proposing a more realistic extension in a remark below. Consider

$$\Delta f(k+1) = \alpha_1 \Delta f(k) + \beta'_1 \Delta P(k),$$

where  $\Delta f(k) = f(k) - f_0$ , and  $f_0$  represents the nominal value of the network frequency, and f(k) the value of the frequency at time k, and where

$$\Delta P(k) = PNx(k) - P_0 = PN(x(k) - x_0),$$

with  $P_0$  representing the power output of the population at time k = 0. We obtain

$$f(k+1) - f_0 = \alpha_1(f(k) - f_0) + \beta'_1 PN(x(k) - x_0).$$

Introducing terms  $\beta_1 = \beta'_1 PN$ ,  $\gamma = (1 - \alpha_1)f_0 - \beta_1 x_0$ , we get

$$f(k+1) = \alpha_1 f(k) + \beta_1 x(k) + \gamma.$$
(5)

Embedding the frequency description in Equation (5) within the dynamics of the Markov chain with n waiting states (2) results in

$$\begin{cases} f(k+1) = \alpha_1 f(k) + \beta_1 x(k) + \gamma \\ x(k+1) = (1-a(k))x(k) + b(k) \sum_{i=1}^n \tau_i w_i(k) \\ w_1(k+1) = b(k) [1-x(k) - \sum_{i=1}^n w_i(k)] \\ w_i(k+1) = b(k)(1-\tau_{i-1})w_{i-1}(k) \\ w_n(k+1) = b(k)[(1-\tau_{n-1})w_{n-1}(k) + (1-\tau_n)w_n(k)], \end{cases}$$
(6)

and into the Markov chain with 3 states (4) to obtain

$$\begin{cases} f(k+1) = \alpha_1 f(k) + \beta_1 x(k) + \gamma \\ x(k+1) = (-k_1 f(k) + k_2) x(k) + (-k_3 f(k) + k_4) \varepsilon_k y(k) \\ y(k+1) = (-k_3 f(k) + k_4) (1 - x(k) - \varepsilon_k y(k)) \end{cases}$$
(7)

#### Dynamical analysis of closed-loop model

We study the stability of the model in (7) with techniques that come from control theory [10], [11]. We are interested in the stability of the model, in the sense that we want to understand the asymptotics of the model, and under which conditions the network frequency will remain within certain (safe) operational bounds. In particular, from Lyapunov stability theory we know that if the Jacobian of a nonlinear system has stable eigenvalues (i.e. absolute value less than 1), then we can deduce the asymptotic stability of the system.

Besides stability, we investigate what characteristics the panels population distributions of the frequency working thresholds must have to lead to disturbances rejection. More precisely, in an actual setup, if the grid frequency goes below a certain safety threshold, a black out is forced by the network operator in order to avoid severe damage to the network itself. The intuition leads to believe that robustness is associated with high variance of population distributions: in particular, dispersed values of disconnection thresholds mean that less panels will disconnect from a given external disturbance.

Let us compute the Jacobian of the vector field in (7), which is a matrix formed by its partial derivatives, as

$$J(f, x, y) = \begin{bmatrix} \alpha_1 & \beta_1 & 0\\ -k_1 x - k_3 \varepsilon_k y & -k_1 f + k_2 & \varepsilon_k (-k_3 f + k_4)\\ -k_3 (1 - x - \varepsilon_k y) & k_3 f - k_4 & \varepsilon_k (k_3 f - k_4) \end{bmatrix}.$$
 (8)

Its determinant and characteristic polynomial may be computed analytically, however with a nontrivial algebraic expression. As such, in order to obtain insight, we consider an identical distribution for a(k) and b(k), which leads to parameters  $k_1 = k_3$  and  $k_2 = k_4$ . This is a reasonable assumption if we consider the semantics of these distributions: they describe the probability to switch ON or OFF, which happens whenever the network frequency is greater or less than a threshold. In practical terms, this means that the threshold related to the ON switch on is the same as that of the OFF switch. The overall model becomes

$$\begin{cases} f(k+1) = \alpha_1 f(k) + \beta_1 x(k) + \gamma \\ x(k+1) = (-k_1 f(k) + k_2)(x(k) + \varepsilon_k y(k)) \\ y(k+1) = (-k_1 f(k) + k_2)(1 - x(k) - \varepsilon_k y(k)) \end{cases},$$

which admits two equilibrium points  $(f_{1,2}^E, x_{1,2}^E, y_{1,2}^E)$ . The Jacobian is

$$J(f, x, y) = \begin{bmatrix} \alpha_1 & \beta_1 & 0\\ -k_1(x - \varepsilon y) & -k_1 f + k_2 & (-k_1 f + k_2)\varepsilon\\ -k_1(1 - x - \varepsilon y) & k_1 f - k_2 & (k_1 f - k_2)\varepsilon \end{bmatrix},$$
(9)

and its associated characteristic 3-rd order polynomial is

$$z^{3} + ((k_{1}f^{*} + k_{2})(1 - \varepsilon) - \alpha_{1})z^{2} + [\beta_{1}k_{1}(x^{*} + \varepsilon y^{*}) + \alpha_{1}(k_{1}f^{*} + k_{2})(\varepsilon - 1)]z - \beta_{1}\varepsilon k_{1}(k_{1}f^{*} - k_{2})$$

We now set conditions on parameters  $k_1$  and  $k_2$  to study the attractivity of the equilibria. Leveraging Rouché arguments [12], we can synthesise the following sufficient condition on the values of  $k_1$  and  $k_2$  where asymptotic stability is guaranteed:

$$\begin{aligned} |(k_1f^* + k_2)(1 - \varepsilon) - \alpha_1| + |\beta_1k_1(x^* + \varepsilon y^*) + \alpha_1(k_1f^* + k_2)(\varepsilon - 1)| + \\ + |-\beta_1\varepsilon k_1(k_1f^* - k_2)| < 1. \end{aligned}$$

In order to practically reason on this condition, we need to define at least some of the numerical values of its unknowns: in the following section we provide approximate values to the unknowns, and accordingly manage to draw conclusions on the stability of the characteristic polynomial.

*Remark 1.* A more realistic transfer function for the grid frequency would be a second order model, namely

$$f(k) = a_1 f(k-1) + a_2 f(k-2) + b_1 x(k-1) + b_2 x(k-2).$$

In order to obtain this second-order model, we have referred to [13], and developed a simple model reproduce the frequency response, taking into account the inertia of the system, the self-regulation of the load, and the primary regulation. This was then discretised in time in order to be compatible with the current framework. The stability structure and the following extended dynamical analysis can be carried out in a similar way. For the sake of brevity and clarity we will stick to the first order model in the following analysis, while in the simulation we will use the second-order dynamics.  $\Box$ 

#### Extension of dynamical analysis to the entire state space

The analysis above holds as long as the frequency remains within certain bounds, namely where the linear approximation of the cdf holds; we need also to take into account other configurations of the system, when the frequency is outside the interval and the evolution of the system changes. Given the switching nature of the system, a hybrid system setup is a rather natural framework. The linearisation of the *cdf* is in fact defined within an interval of frequencies, outside of which a(k) and b(k) have a steady value of 0 or 1. We study those cases in the following.

We focus our attention on the system with five configurations. Under the assumption of the distributions of a(k) and b(k) to be identical, we argue that we model a hybrid system with 5 different configurations; in case of different distributions, the number of configurations may increase, so we do not study them in full in this paper for brevity. Figure 4 shows the divisions into the five different configurations.

Our interest is to analyse the connection between the various configurations of the system and to understand the conditions to keep the frequency close to the nominal value and to avoid its drift to zero.



Fig. 4: Schematic diagram of the 5 frequency configurations, indicated as circled numbers.

1. Configuration 1: this configuration describes the case where the network frequency belongs to the set of values closest to the nominal value. Our system becomes an LTV system, given the conditions  $1 - a(k) \equiv 1 \equiv b(k)$ ,  $f(k) \in [-\nu, \nu]$ :

$$\begin{cases} f(k+1) = \alpha_1 f_k + \beta_1 x(k) + \gamma \\ x(k+1) = 1 \cdot (x(k) + \varepsilon_k y(k)) \\ y(k+1) = 1 \cdot (1 - x(k) - \varepsilon_k y(k)) \end{cases}$$
(10)

There is only one equilibrium point,  $p^E = \left(\frac{\beta_1}{1-\alpha_1}(1-x_0) + f_0, 1, 0\right)$ . From the analysis of the evolution matrix, the equilibrium is asymptotically stable. If the contribution coming from  $\frac{\beta_1}{1-\alpha_1}(1-x_0)$  is small enough, the equilibrium remains inside the interval  $[f_0 - \nu, f_0 + \nu]$ , whereas if the contribution is too large the system will switch to Configuration 2.

2. Configuration 2: This is the case where the evolution matrix of the system becomes non linear and time varying, with  $f(k) \in [\nu, \nu + \eta]$ . The system may be written as

$$\begin{cases} f(k+1) = \alpha_1 f(k) + \beta_1 x(k) + \gamma \\ x(k+1) = (1-a(k))x(k) + b(k)\varepsilon_k y(k) \\ y(k+1) = b(k)(1-x(k)) - b(k)\varepsilon_k y(k) \end{cases}$$

and as we stated, we assume the same distribution for a(k) and b(k), which we explicit  $1 - a(k) = -k_1 f(k) + k_2 = b(k)$  where  $k_1, k_2 > 0$ . To find the equilibria we set that

$$f = \frac{\beta_1}{1 - \alpha_1} x + \frac{\gamma}{1 - \alpha_1}, \quad x + y = -k_1 f + k_2,$$

and from these two equations we obtain

$$y = \left(\frac{-k_1\beta_1}{1-\alpha_1} - 1\right)x - \frac{k_1}{1-\alpha_1}\gamma + k_2.$$

Substituting this value into the previous equation, we have

$$\frac{-k_1\beta_1}{1-\alpha_1} \left(1+\varepsilon \left(\frac{-k_1\beta_1}{1-\alpha_1}-1\right)\right) x^2 + \left[\left(-\frac{k_1}{1-\alpha_1}\gamma+k_2\right) \left(1-\varepsilon+2\varepsilon \left(\frac{-k_1\beta_1}{1-\alpha_1}\right)\right)-1\right] x+\varepsilon \left(-\frac{k_1}{1-\alpha_1}\gamma+k_2\right)^2\right]$$

The solution of this 2-nd order equation can be found symbolically, however in order to garner insight on its solutions, we set some indicative (and practically meaningful) values for the variables ranges:

$$k_1 \sim 10, k_2 \sim 10^2, \alpha_1 \sim 1 - \alpha_1 \sim 10^{-1}, \varepsilon \sim 10^{-1}, \beta_1 \sim N,$$

where N represents the number of solar panels, assumed to be  $N \gg 1$ . This results in equation  $10^3 N^2 x^2 - 2 \cdot 10^3 N x + 10^3$ , which is endowed with two coincident solutions,  $x_{1,2}^E = \frac{1}{N}$ . This results in the following approximated equilibrium point:

$$p_{1,2}^E \simeq \left(\frac{1+\gamma}{1-\alpha_1}, x_{1,2}^E, 0\right),$$

where the value of x is less than 1, giving a frequency to be *less* than the value we found in Configuration 1. The equilibrium could, in this case, be inside the thresholds of Configuration 1 or remain inside Configuration 2 depending on the numerical values of the several parameters; we will discuss this in Appendix B.

We state that the equilibrium is attractive because we argue that the system in Configuration 1 presents variables f, x, y which are always greater than or equal to the ones in this configuration: a(k) and  $b(k) \in [0, 1] \forall k$ . Since the stability is asymptotic in the first case, we conclude the asymptotic stability also in this configuration.

- 3. Configuration 3: this configuration describes the behaviour of the system when it is completely outside the range of the frequency thresholds, i.e. 1 a(k) = 0 = b(k). The system becomes null, namely x(k + 1) = 0, y(k + 1) = 0, the frequency only evolves with a decreasing exponential, i.e. the frequency will move towards the cases discussed above.
- 4. **Configuration 4**: the analysis is similar to the one of Configuration 2. In this case it is important to notice that, since the frequency at the equilibrium point is less than the nominal value, the system could move towards Configuration 1, remain in the current configuration or slide towards a decreasing frequency: this depends on the numerical value of the parameters.
- 5. **Configuration 5**: the analysis is analogous to Configuration 3. The frequency will drop to zero.

# 4 Experimental evaluation of the aggregated models

In order to show the precision of the aggregated models, we set up rounds of simulations comparing the two alternative models (that with n waiting states in Fig. 2 and the one with three locations in Fig. 3) with the ground truth obtained from an explicit simulation of the entire population of PV panels within the power network (which we denote as the *explicit* model).

For the *explicit* model, each of the N panels has been given four different frequency thresholds (disconnection and reconnection in over- and under-frequency) and a time delay (number of time steps the devices need to wait before turning to

ON). These parameters have been generated according to set probability distributions (see below) for the population, which are then used in the computations for the abstract models. We have set up the distributions of frequency thresholds as uniform, and that for the time delay as geometric, and we set  $N = 10^6$ . The assumption of a uniform pdf allows us to exactly define the variables  $\nu$ ,  $\eta$  mentioned in the previous section, where more generally we utilised a linear approximation of the cdf, e.g.  $b(k) = cdf^{1|0}(f(k)) \simeq -k_1 f(k) + k_2$ . For simplicity,  $P_i$  was set to be constant and equal to  $\bar{p} = 3 \ kW$  for each device. With reference to the discussion in Remark 1, the frequency dynamics have been described by a second-order difference equation, as

$$f(k+1) = \alpha_1 f(k) + \alpha_2 f(k-2) + \beta_1 x(k) + \beta_2 x(k-1),$$

where the constants  $\alpha_1, \alpha_2, \beta_1, \beta_2$  are set to make the transfer function stable. Our two aggregated models are derived from the following simplifications:

- 1. Frequency thresholds: we have an exact description of the frequency thresholds and a linear dependency from f(k) in the extended system;
- 2. Time delays: we have introduced 1 s time delays and defined waiting states accordingly. We have defined a maximum possible delay as n, and in order to cope with delays longer than n, we have set a self loop on the n-th state;
- 3. Lumped delays: we have lumped together the *n* waiting states, introducing an approximation encoded in  $\varepsilon_k$ .

As motivated in the Introduction to this work, we have set up two specific simulation scenarios:

- 1. panels disconnecting in view of an external disturbance;
- 2. effect of the distribution of frequency disconnection/reconnection thresholds, in response to an external disturbance.

**First scenario** We set up a stable network with N devices initially in the ON mode, when the grid relies on their power production to be stable and to guarantee a reliable service to the load. The initial condition is set to  $x_0 = 1$ , and the grid frequency is set to the stable value,  $f_0 = 50 Hz$ . At time t = 10s we inject an external disturbance, in order to create a frequency peak of 50.16 Hz, and to observe the ensuing dynamics in the network.

The comparison among the three models is in Figure 5 and in Figure 6, in terms of network frequency response and portion of ON panels. On the one hand, we observe from Figure 5 that both the two abstract models seem to be a lowpass filtered signal of the explicit one. The model with n delay states follows the dynamics of the explicit one, and the model with one delay state follows that with n states, thanks to the estimation of the quantity  $\varepsilon_k$  (as discussed in Appendix A). On the other hand, looking at Figure 5 we note that the difference in terms of percentage of active panels is always less than 2%. This difference is due to the low-pass action of the abstract models, which leads to slower convergence to the same equilibrium point as the explicit model. These figures show that our abstract models can reflect the evolution of the explicit model in a reliable way.



Fig. 5: Comparison of network frequency behaviour (left) and of ON population (right) for the explicit model (blue, circles), the abstract model with n delay states (red, crosses) and the abstract model with 3 states (green, triangles).

Second scenario We run simulations to test if more variance in the reconnection/disconnection thresholds brings a more stable system, as intuition suggests. Figure 6 shows that, as expected, a higher variance is desirable for the model under consideration. In the explicit system, at time t = 10 s an external disturbance was injected in order to make the frequency jump to 50.16 Hz, value for which a portion of devices will disconnect. Simulations run with 10 different values of variance (v) for the uniform distributions, resulting in bigger oscillations and longer time to get frequency back to the nominal value for small v; values of v close to 5 makes the system almost insensible to the disturbance.



Fig. 6: Comparison of network frequency behaviour with different degrees of variance in the frequency thresholds: smaller variance produces bigger oscillations. Simulations run with the explicit model.

We have finally run simulations to display the effect of several frequency evolution parameters in the Appendix B, where we show how the different position of the equilibria, as discussed in Section 3, affect the response of the system.

### 5 Conclusions

We have presented a modelling framework for the aggregation of a population of photovoltaic panels, and studied its dynamics when interacting within the electric grid and its frequency behaviour. We have provided experiments comparing two models against a ground truth, where N photovoltaic panels were singularly simulated: a scenario has shown how reliably the models behave in presence of a massive network failure, another how population heterogeneity influences the performances of the system.

Current research emphasis concerns variable power outputs, (modelling uncertainties such as weather, occlusions), and novel global control schemes.

### References

- Solar Power Europe. Global Market Outlook for Solar Power 2016-2020. Technical report, 2016.
- K. Ellis H. Tiam, F. Mancilla-David. A Detailed Performance Model for Photovoltaic Systems. Technical report, NREL, 2012.
- 3. H. Wirth. Recent Facts about Photovoltaics in Germany. Technical report, Fraunhofer ISE, 2016.
- 4. Dispersed Generation Impact on CE Region, Dynamic study. Technical report, entsoe, 2014.
- Regional Group of Continental Europe and Synchronous Area Great Britain. Solar Eclipse 2015 - Impact Analysis -. Technical report, entsoe, 2015.
- B. Lazpita M. Jung, O. Wiss. Analyses et conclusions "Tests en sous-frequence". Technical report, RTE, 2016.
- B. Lazpita M. Jung, O. Wiss. Analyses et conclusions "Tests en sur-frequence". Technical report, RTE, 2016.
- Legal Comparison of PV Regulations. http://www.res-legal.eu/comparison-tool/. Accessed: 04/04/2017.
- 9. J. K. Moller E. B. Iversen, J. M. Morales. Probabilistic Forecasts of Solar Irradiance by Stochastic Differential Equations. *Envirometrics*, 25, April 2014.
- 10. A. J. Krener R. Hermann. Nonlinear Controllability and Observability. *IEEE Transaction on Automatic Control*, 1977.
- 11. H. K. Khalil. Nonlinear Systems. Prentice Hall, 2002.
- A. F. Beardon. Complex Analysis: The Argument Principle in Analysis and Topology. 1979.
- 13. Policy1: Load-Frequency Control and Performance. Technical report, enstsoe, 2009.

### A Observability and observer design

In a practical instance, whilst it is in principle possible to have real-time information about every photovoltaic panel in the network, economic reasons and the sheer size of the modelling problem would make this impractical. We thus consider a more realistic scenario, where the only observable output is the frequency of the network.

#### Model observability

Let us prove the systems, both the one with 3 states in the Markov Chain and the one with n waiting states are in fact observable when the only output signal is f(k). This means, in practical terms, that we can infer the value of the other variables of the system, reconstructing them solely from the value of f(k). We focus on the abstract one firstly.

The output matrix is simply the constant

$$C(k) \equiv C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

and consider the matrix of the linearized system

$$A = \begin{bmatrix} \alpha_1 & \beta_1 & 0\\ -k_1x - k_3\varepsilon y & -k_1f + k_2 \varepsilon(-k_3f + k_4)\\ -k_3(1 - x - \varepsilon y) & k_3f - k_4 & \varepsilon(k_3f - k_4) \end{bmatrix}$$

To simplify the computations consider  $k_1 = k_3 := k$  and  $k_2 = 0 = k_4$ . This gives an observability matrix

$$\mathcal{O} = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_1 & \beta_1 & 0 \\ \alpha_1^2 - k_1 \beta_1 (x + \varepsilon y) & \beta_1 (\alpha_1 - k_1 f + k_2) & \beta_1 \varepsilon (-k_1 f + k_2) \end{bmatrix}.$$

The rank is full as long as  $f \neq k_2/k_1$ : this condition holds within the range where the linear approximation holds.

A similar conclusion can be drawn for the bigger system, in which each  $w_i(k)$  is dependent from  $w_{i-1}(k-1)$ . This can be thought as a chain of substitutions that guarantees the observability in n steps.

#### **Observer** design

An observer can in fact be built for both models. We start with the observer for the smaller one. Assume the output is just the frequency of the network, namely f(k) and we want to obtain the state x(k). In this case we may think about  $\varepsilon_k = \tau_1 \forall k$  to make computation straightforward, or to think to know how the value of  $\varepsilon_k$  is computed at each time step. From the dynamical equations one easily obtains

$$\begin{bmatrix} \hat{f}(k-2)\\ \hat{x}(k-2)\\ \hat{y}(k-2) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1\\ 0 & 1/\beta_1 & -\alpha_1/\beta_1\\ \frac{1}{\beta_1 b(k-2)\varepsilon_{k-2}} & -\frac{(\alpha_1+1-a(k-2))}{\beta_1 b(k-2)\varepsilon_{k-2}} & \frac{\alpha_1(1-a(k-2))}{\beta_1 b(k-2)\varepsilon_{k-2}} \end{bmatrix} \times \\ \times \begin{bmatrix} f(k)\\ f(k-1)\\ f(k-2) \end{bmatrix}.$$

Our estimation is two steps behind the current time instant. We then set up a two-steps predictor following the dynamical equations of the system, after some algebraic calculi, as

$$\begin{bmatrix} \hat{f}(k) \\ \hat{x}(k) \\ \hat{y}(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ obs_{2,1} & obs_{2,2} & obs_{2,3} \\ obs_{3,1} & obs_{3,2} & obs_{3,3} \end{bmatrix} \begin{bmatrix} f(k) \\ f(k-1) \\ f(k-2) \end{bmatrix} + \\ + \begin{bmatrix} 0 \\ b(k-1)\varepsilon_{k-2}b(k-2) \\ b(k-1)(1-\varepsilon_{k-2}b(k-2)) \end{bmatrix},$$

where

$$obs_{2,1} = \frac{1 - a(k-1) - b_{k-1}\varepsilon_{k-1}}{\beta_1},$$
  

$$obs_{2,2} = \frac{b(k-1)\varepsilon_{k-1}(1 - a_{k-2} - b_{k-2} + \alpha_1) - \alpha_1(1 - a_{k-2})}{\beta_1},$$
  

$$obs_{2,3} = \frac{b(k-1)\varepsilon_{k-1}\alpha_1(b_{k-2} - (1 - a_{k-2}))}{\beta_1},$$
  

$$obs_{3,1} = -\frac{b(k-1)}{\beta_1}(1 + \varepsilon_{k-1}),$$
  

$$obs_{3,2} = \frac{b(k-1)}{\beta_1}(\alpha_1 - \varepsilon_{k-1}(\alpha_1 + 1 - a_{k-2} - b_{k-2})),$$
  

$$obs_{3,3} = \frac{b(k-1)}{\beta_1}\varepsilon_{k-1}\alpha_1(b_{k-2} + a_{k-2} - 1).$$

For the system with n waiting states a similar situation is expected. In fact, for the dynamical equation we need an n + 2 steps predictor. The key part is how to connect the estimation of x(k) to  $w_1(k)$ . We note that

$$x(k+1) + \sum_{i} w_i(k+1) = (1 - a(k))x(k) + b(k)(1 - x(k)),$$

so that

$$\sum_{i} w_i(k) = (1 - a(k - 1))x(k - 1) + b(k - 1)(1 - x(k - 1)) - x(k).$$

We can substitute this into the estimator of  $w_1(k)$ , giving

$$\hat{w}_1(k) = b(k-1)[1 - b(k-1)(1 - \hat{x}(k-1)) - (1 - a(k-1))\hat{x}(k-1)].$$

The observer can be built in a recursive way, as

$$\begin{cases} \hat{f}(k) = f(k) \\ \hat{x}(k-1) = \frac{f(k) - \alpha_1 f(k-1)}{\beta_1} \\ \hat{x}(k) = (1 - a(k-1))\hat{x}(k-1) + b(k-1)\sum_i \tau_i \hat{w}_i(k-1) \\ \hat{w}_1(k) = b(k)[1 - b(k-2)(1 - \hat{x}(k-2)) - (1 - a(k-2))\hat{x}(k-2)] \\ \hat{w}_i(k) = b(k-1)(1 - \tau_{i-1})\hat{w}_{i-1}(k-1) \quad for \ i = 2, \dots n-1 \\ \hat{w}_n(k) = b(k-1)[(1 - \tau_{n-1})\hat{w}_{n-1}(k-1) + (1 - \tau_n)\hat{w}_n(k-1)] \end{cases}$$

With this technique we proved that we are able to build an observer for the system, estimate the  $w_i(k)$  and compute the  $\hat{\varepsilon}_k$  value at each time step as

$$\hat{\varepsilon}_k = \frac{\sum_i \tau_i \hat{w}_i(k)}{\sum_i \hat{w}_i(k)},$$

where  $\hat{w}_i(k)$  is the estimate of  $w_i(k)$ . The observer has a transitory of (n+2) step, necessary to compute all the  $w_i$  initially.

## **B** Simulations for stability analysis

We have developed more experiments in order to show the importance of the initial conditions for the overall dynamics. In Section 2 we have divided the behaviour of the model into five configurations, in order to study its steady state and to understand whether the network is stable or if it admits unstable equilibria. We have defined Configuration 1 as the one where the network frequency is around its nominal value, namely  $f(\cdot) \in [f_0 - \nu, f_0 + \nu]$ , and such that a(k) = 0, b(k) = 1, forcing all panels to be turned to the ON state in the long run. We have found that the equilibrium for the frequency is

$$f = \frac{\beta_1}{1 - \alpha_1} (1 - x_0) + f_0,$$

where  $f_0$  is the nominal value of the frequency,  $x_0$  is the initial portion of panels turned ON, and  $\frac{\beta_1}{1-\alpha_1}$  depends on the evolution equation of the frequency. If  $\frac{\beta_1}{1-\alpha_1}(1-x_0) \leq \nu$ , the equilibrium point is inside Configuration 1, otherwise the equilibrium is inside the interval for Configuration 2. This translates into the situation where the network can or cannot handle the global population of PV, leading to a stable or an unstable situation, respectively. In order to be able to understand the effect of the frequency evolution on the overall closed-loop model, we set up additional rounds of simulations. For simplicity of notation let us define  $\gamma' := \frac{\beta_1}{1-\alpha_1}$ . We imagine a null portion of panels in the ON state initially, namely  $x_0 = 0$ , an initial frequency at  $f_0$  and different values for the term  $\gamma'$ : the latter choice allows one to to move the equilibrium point of Configuration 1 inside or outside the range  $[f_0 - \nu, f_0 + \nu]$ . We keep constant values for  $\nu$  and  $\eta$ . Three situation are tested, with  $\gamma' = 1$ ,  $\gamma' = 0.5$  and  $\gamma' = 0.1$ . In the first two cases, the equilibrium resides outside the limits of Configuration 1, while in the last case the equilibrium is inside this interval. As shown in Figure 7, we obtain three different equilibria corresponding to different portions of the ON population.

The first instance,  $\gamma' = 1$ , is the more unstable one. PV panels, after the delay of around 20 s try to turn ON. This causes a jump in frequency that moves the system into Configuration 2, where we see an oscillatory behaviour before reaching the stable point at x = 0.27 with a frequency f = 50.27 Hz (not shown). The equilibrium is in fact inside the limits of Configuration 2.

The second case, with  $\gamma' = 0.5$ , shows a less oscillatory behaviour while the system slides from Configuration 1 to Configuration 2. In this case the stable point is set at x = 0.47 and f = 50.24 Hz (not shown).

The last instance, when  $\gamma' = 0.1$ , shows an equilibrium inside Configuration 1. The system is free to turn to ON the whole population of PV panels with a smooth evolution in frequency (not shown). The equilibrium in this case is x = 1 and  $f = 50.1 \ Hz$ .



Fig. 7: Portion of panels ON under different frequency evolutions.