

# Model of hematopoiesis dynamics under IFN $\alpha$ therapy in Myeloproliferative Neoplasms

**Abstract.** Hematopoiesis is a complex process in which stem cells in the bone marrow produce all the cells circulating in the blood. When specific mutations occur in the stem cells, such as the  $JAK2^{V617F}$  mutation, for example, this process can be altered leading to the development of hematological malignancies such as Myeloproliferative Neoplasms. Interferon alpha (IFN $\alpha$ ) is a treatment that allows a hematological response in patients, but also in some cases a molecular response. However, its precise mechanism of action is still poorly understood, and there are no clear guidelines on the use of this treatment for clinicians. In this article, we model the action of the IFN $\alpha$  on the hematopoiesis dynamics in order to study in particular its effect at the level of stem cells. Using a Bayesian parameter estimation method and data from a patient under treatment for several years, we show that IFN $\alpha$  could act on cancer stem cells by promoting their quiescence exit and thus their proliferation while allowing their exhaustion from the stem cells stock by increasing their propensity to make differentiated divisions.

**Keywords:** Dynamic Systems · Computational Bayesian Statistics · CMA-ES · Hematopoiesis · IFN $\alpha$  Therapy · Myeloproliferative Neoplasms

## 1 Introduction

Hematopoiesis is the process leading to the production of all blood cells. This is a hierarchical system at the top of which is the hematopoietic stem cell (HSC). HSCs harbor two properties: long-term self-renewal and multipotency (property to give rise to all blood cell types). HSCs can commit to successive types of progenitors, which are cells engaged in cell differentiation and which retain the potential to give several types of blood cells. More restricted progenitors commit to precursors which lose the ability to proliferate but can mature into fully differentiated cells. Hematopoiesis is, therefore, a succession of cell differentiation and proliferation steps involving many different cell types. HSC deregulation can give rise to hematological malignancies. Among them, Myeloproliferative Neoplasms (MPN) occur as a result of the mutation of specific genes in HSC located in the bone marrow. This leads to a disturbance in hematopoiesis. In the particular case of Polycythemia Vera (PV), it results in an overproduction of red blood cells in the patient which can lead to medical complications such as thromboses or hemorrhages. These pathologies can also progress to secondary leukemia. Understanding MPN diseases remains a challenge for biologists, both

because hematopoiesis is a complex dynamic system and because HSCs are difficult to access. Several driver mutations for MPN have already been identified, including the  $JAK2^{V617F}$  gain-of-function mutation in the gene that codes for Janus Kinase 2, a protein that plays a crucial role in cell signalling [1]. However, the precise role of this mutation in the dynamics of the disease and the selective advantage it provides during all steps of cell differentiation remains incompletely understood.

Advances in the understanding of this disease are crucial to enable the development of treatments that will lead to a patient’s recovery. Interferon alpha ( $IFN\alpha$ ), a natural inflammatory cytokine that has long been used in the treatment of many diseases, has shown promising results in MPN. Indeed,  $IFN\alpha$  induces not only a hematological response, i.e. a normalization in blood cell counts, but also a molecular response, i.e. a reduction of mutated cells [2, 3]. In order to enable the development of personalized medicine, it is necessary to be able to understand and quantify the impact of this treatment on (mutated) hematopoietic cells, particularly in HSC.

Mathematical modelling and proper statistical methods can help biologists and physicians achieve these goals. There are many proposals for mathematical models of hematopoiesis in the scientific literature [4–6]. In particular, they have been widely used to describe leukaemia, as in [7, 8] for example. One of the advantages of these models and their usefulness to biologists is the possibility of inferring stem cell dynamics. For example, Michor et al. [9] show with a compartmental model that imatinib targets leukemic stem cells which can lead to a molecular remission.

In this paper, we propose a new compartmental model of hematopoiesis to describe the dynamics of healthy and cancer hematopoietic cells under the action of  $IFN\alpha$ . In particular, we focus on the drug effect on the HSC. We make the hypothesis that the treatment can lead to a molecular remission by promoting quiescence exit and differentiated divisions of the HSC. We develop a method for estimating the parameters of the model based on a Bayesian framework. Finally, we apply this method to data from a PV patient undergoing treatment to validate our hypotheses.

## 2 Model

### 2.1 Data and experimental setup

**Observational patient study.** At Gustave Roussy cancer research institute, an observational prospective study is underway on a cohort of 50 MPN patients treated with  $IFN\alpha$  for several years. These patients are mainly suffering from Essential Thrombocythemia or Polycythemia vera (PV), and harbor mutations in  $JAK2$ ,  $CALR$  or  $MPL$  gene [1]. The patients generally present a mix of cells that are heterozygous (denoted with the symbol ‘het’, mutation on one allele), homozygous (‘hom’, mutation on both alleles following a mitotic recombination) and even healthy that are called wild-type (‘wt’). The patient considered in this article is a 59-year-old male individual with a  $JAK2^{V617F}$  PV. Written

informed consent was obtained in accordance with the Declaration of Helsinki and the study was approved by the Ethics Committee from CHU DIJON and Commission Nationale de l'Informatique et des Libertés (CNIL) (authorization #915663). He had been under treatment with IFN $\alpha$  for six years and received a variable posology with primarily high doses of IFN $\alpha$ , up to 180 $\mu$ g weekly. Most of his mutated cells are homozygous. Data were collected approximately every four months for four years and annually afterwards. We denote  $\{t_i\}_{i \in I}$  the set of sampling dates, relatively to the start of treatment, with I indexing the measurements.

**Measures.** Using regular blood samples, we measured the proportions of mutated cells at two levels. First, at the level of mature cells, we measured for the granulocytes the global allelic fraction of mutated genes by Taqman allelic discrimination qPCR. That is, for these cells, we cannot distinguish between het and hom cells. For all  $i \in I$ ,  $\hat{y}_i$  is the allelic fraction measured at time  $t_i$ .

The second level involves the earliest stages of hematopoietic dynamics corresponding to immature cells. Based on the surface markers CD34<sup>+</sup>CD38<sup>+</sup>, CD34<sup>+</sup>CD38<sup>-</sup>CD90<sup>-</sup> and CD34<sup>+</sup>CD38<sup>-</sup>CD90<sup>+</sup>, we isolated the different progenitor cells from the others cells, purified them and sorted at one progenitor per well in 96-well plates. Each progenitor gave a progeny of cells (colony) after 10 to 15 days of culture. The genotyping of each colony by Taqman allelic discrimination qPCR enabled to retrospectively know the genotype of each progenitor. This allowed us to know precisely how many wt, het and hom immature cells were in our subsample. At a given time  $t_i$  for  $i \in I$ , we denote these numbers  $\hat{n}_i, \hat{n}'_i$  and  $\hat{n}''_i$  respectively. Because this subsample is of reduced size (around 200 progenitors), the mutated immature cell proportions deduced are not fully representative of those from the entire progenitor population within the patient's body. This induces a sampling noise that we model adequately in the next paragraph.

## 2.2 Noise model

**Mature cells.** At time  $t_i$ , the measured clonal fraction  $\hat{y}_i$  corresponds to an unknown real fraction  $y_i$ . We have  $\hat{y}_i, y_i \in [0, 1]$ . Conventional uncertainty models are additive or multiplicative Gaussian noises that are not entirely appropriate in our case. We choose to generalize them as follows:

$$\hat{y}_i | y_i \sim \mathcal{N}(y_i, \sigma^2(y_i)) \quad (1)$$

with  $\sigma^2 : [0, 1] \rightarrow \mathbb{R}$ . By choosing for the function  $\sigma^2$  a constant, we find ourselves with an additive noise and by choosing  $\sigma^2(y_i) = \sigma_m^2 y_i^2$  we find ourselves with a multiplicative noise. In this article, in order to have a noise that is maximum at  $y_i = 1/2$  and vanishes for  $y_i = 0 = 1$ , we choose:

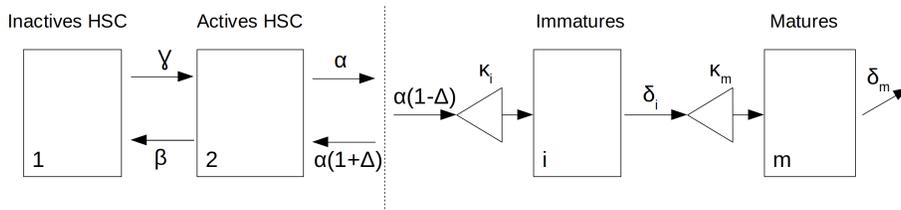
$$\sigma^2(y_i) = y_i(1 - y_i)\sigma_m^2 \quad (2)$$

**Immature cells.** To model the sampling noise we have for the mutated clonal fractions of progenitors, we assume that we have randomly drawn with replacement immature cells from the patient’s body. This approach is used, for example, by Catlin et al. [10] and allows to model the uncertainty by a multinomial law. For a large number of immature cells in the body, the approach is almost identical to that of Xu et al. [5] who consider a multivariate hypergeometrical law used to model sampling without replacement. Let’s consider that at date  $t_i$  for  $i \in \mathbb{I}$ , the real proportions of het and hom immature cells are respectively  $z'_i$  and  $z''_i$  (and  $z_i = 1 - z'_i - z''_i$ ). From the set of immature cells, of unknown but very large number, we draw a number  $\hat{N}_i := \hat{n}'_i + \hat{n}''_i + \hat{n}_i$  of cells. Among these cells, we have exactly  $\hat{n}'_i$  het and  $\hat{n}''_i$  hom mutated cells. Since they are random variables, they follow a multinomial law:

$$\mathbb{P}[\hat{n}_i = n_1, \hat{n}'_i = n_2, \hat{n}''_i = n_3 | z'_i, z''_i] = \frac{(n_1 + n_2 + n_3)!}{n_1!n_2!n_3!} z_i^{n_1} z'_i{}^{n_2} z''_i{}^{n_3} \quad (3)$$

### 2.3 Compartmental model

**One brick for one cellular type.** To describe the dynamics of cell differentiation and proliferation, we propose a compartmental model in which each compartment represents a category of cells in the hematopoietic hierarchical structure. This type of model is commonly used in the literature, for example, in [7, 9]. Our first assumption is the absence of interaction between wt, het and hom mutated cells. We choose to treat these three types independently and identically. For a given type, we will then end up with a system of Ordinary Differential Equations (ODE), with parameter values that are specific to the cell type under consideration. This is what we call the basic building brick of the model, shown in Figure 1. Two compartments correspond to the non-stem



**Fig. 1.** One brick of the model. Schema of the 4 hematopoietic compartments for a given cellular type (wild-type, heterozygous or homozygous) and under given treatment conditions (with or without IFN $\alpha$ ) and the associated parameters.

cells, related to the data in our possession (see § 2.1). Mature cells will die at the rate  $\delta_m$ . Immature cells leave their compartment at the rate  $\delta_i$ . The parameter  $\kappa_m$  models the proliferation of cells from progenitors to mature cells. Two compartments correspond to stem cells. Indeed, HSCs are far from being

a homogeneous whole. Instead, there appears to be a continuum of states, from proliferating active stem cells to quiescent states. Compartment 2 represents active cells, which can proliferate. These cells are recruited at a rate  $\alpha$  to contribute to hematopoiesis. Once recruited, these cells divide. We consider three possible types of division for stem cells [15]:

- Asymmetric division, where one HSC generates one HSC and one differentiated cell, with a frequency of occurrence equal to  $p_1$
- Symmetrical division, where one HSC generates two HSCs, with a frequency of occurrence equal to  $p_2$
- Differentiated division, where one HSC generates two differentiated cells, with a frequency of occurrence equal to  $p_0 = 1 - p_1 - p_2$

We denote  $\Delta := p_2 - p_0$  which expresses the balance between these three mechanisms of division. For  $\Delta = 0$ , we are in the conditions of homeostasis, for  $\Delta < 0$  there is an exhaustion and for  $\Delta > 0$  an invasion of the stem compartments. Finally, the rates  $\gamma$  and  $\beta$  model exchanges between compartments 1 and 2. We do not make any assumption on the nature of these two compartments. In the biological literature, one could find, for example, the distinction between dormant or non-dormant [11]. In the end, the system of equations to describe the dynamics of a given hematopoietic cell population (wt, het or hom) is as follows:

$$\begin{cases} \frac{dN_1(t)}{dt} = -\gamma N_1(t) + \beta N_2(t) \\ \frac{dN_2(t)}{dt} = \gamma N_1(t) + (\alpha\Delta - \beta)N_2(t) \\ \frac{dN_i(t)}{dt} = \alpha(1 - \Delta)\kappa_i N_2(t) - \delta_i N_i(t) \\ \frac{dN_m(t)}{dt} = \delta_i \kappa_m N_i(t) - \delta_m N_m(t) \end{cases} \quad (4)$$

With  $N_1(t)$ ,  $N_2(t)$ ,  $N_i(t)$  and  $N_m(t)$  the numbers of cells, of a given type, in compartments 1, 2, immature and mature respectively at time  $t$ . The proposed ODE system is linear, and we can derive an analytical solution.

**Considering all cell types.** We have described above the dynamics for a population of cells of a given type. In practice and to be consistent with our experimental data, we do not have access to the quantities of wt, het or hom cells separately, but to their relative proportions. To take this into account, we consider that each population of a given type follows an ODE system as described in eq. (4). Then we will consider as outputs of our model no longer the numbers of cells but the proportions of immature heterozygous and homozygous cells respectively  $z'(t) = \frac{N_i'(t)}{N_i(t) + N_i'(t) + N_i''(t)}$  and  $z''(t) = \frac{N_i''(t)}{N_i(t) + N_i'(t) + N_i''(t)}$  as well as the mature cell allele burden  $y(t) = \frac{0.5 \cdot N_m'(t) + N_m''(t)}{N_m(t) + N_m'(t) + N_m''(t)}$  where every contribution of het cells is only counted for one half. In terms of notation, we use zero, one and two apostrophes for wt, het and hom cells. For convenience, we define  $k_i'$  such that  $\kappa_i' = k_i' \kappa_i$  and likewise  $k_i''$ ,  $k_m'$  and  $k_m''$ .

**Effect of IFN $\alpha$  and initial conditions.** So far, we have not talked about the effect of IFN $\alpha$  in the model. Following the idea of Michor et al. [9], we consider that IFN $\alpha$  acts by modifying the values of some parameters in the model. We consider that time  $t = 0$  corresponds to the beginning of the treatment. Before that time, the equations (4) are still valid, but we consider that the homeostatic conditions are satisfied, i.e. the system is in a quasi-stationary state. This is of course verified for wt cells as soon as  $\Delta = 0$ , which gives us the following initial conditions:  $N_1(0) = \frac{\beta}{\beta+\gamma}N_{HSC}$ ,  $N_2(0) = \frac{\gamma}{\beta+\gamma}N_{HSC}$ ,  $N_i(0) = \frac{\kappa_i\alpha}{\delta_i}N_2(0)$  and  $N_m(0) = \frac{\kappa_m\delta_i}{\delta_m}N_i(0)$  with  $N_{HSC}$  the total wild-type HSC number considered constant. For mutated cells, we know that the homeostatic condition is not verified since they tend to invade the stem cell compartment. But since the invasion happens over a long time (over 30 years), we assume  $\Delta' \approx \Delta'' \approx 0^+$ . We introduce  $\eta' = \frac{N'_1(0)+N'_2(0)}{N_{HSC}}$  and  $\chi' = \frac{N'_2(0)}{N'_1(0)+N'_2(0)}$ . We can then express the initial conditions for het cells:  $N'_1(0) = \eta'(1 - \chi')N_{HSC}$ ,  $N'_2(0) = \chi'\eta'N_{HSC}$ ,  $N'_i(0) = \frac{\kappa'_i\alpha'}{\delta'_i}N'_2(0)$  and  $N_m(0) = \frac{\kappa'_m\delta'_i}{\delta'_m}N'_i(0)$ . The same goes for hom cells. From  $t = 0$ , the patient is under treatment. IFN $\alpha$  will then modify the values of the parameters, potentially in different ways depending on the cell type. In terms of notation, we add the subscript  $*$  to the parameters impacted by the drug. We define  $k_{i*}$  such that  $\kappa_{i*} = k_{i*}\kappa_i$  and similarly for  $k'_{i*}$ ,  $k''_{i*}$ ,  $k_{m*}$  and  $k''_{m*}$ . From  $t \geq 0$ , eq. (4) remain valid with new parameters, and there is an equilibrium shift. This induces a new dynamics. From our data, we estimate the values of the model parameters. This is the subject of the next section which presents the parameter estimation method.

### 3 Method

#### 3.1 Bayesian framework and MCMC algorithm

To estimate the parameters of the model described above, we consider a Bayesian framework. Let  $\mathcal{M}$  be the model considered,  $\sigma_m^2$  the parameter introduced in eq. (2) and  $\boldsymbol{\theta}$  the random vector of the parameters to be estimated. Our goal is to estimate the posterior distribution of  $\boldsymbol{\theta}$  knowing the data:

$$\mathbb{P}[\boldsymbol{\theta}, \sigma_m^2 | \mathcal{D}, \mathcal{M}] \propto \mathbb{P}[\mathcal{D} | \boldsymbol{\theta}, \sigma_m^2, \mathcal{M}] \mathbb{P}[\boldsymbol{\theta}] \mathbb{P}[\sigma_m^2] \quad (5)$$

with  $\mathcal{D} = \{\hat{n}'_i, \hat{n}''_i, \hat{n}_i, \hat{y}_i\}_{i \in \mathbb{I}}$  the set of all measures for the considered patient. For clarity, we omit to mention  $\mathcal{M}$  in the following and we consider that  $\sigma_m^2$  is included in the parameter vector  $\boldsymbol{\theta}$  we want to estimate. The prior distribution is chosen uniform (see table 1). The expression of the likelihood is obtained from the noise model presented in §2.2. We have

$$\mathbb{P}[\mathcal{D} | \boldsymbol{\theta}] \stackrel{(1)}{=} \prod_{i \in \mathbb{I}} \mathbb{P}[(\hat{n}'_i, \hat{n}''_i, \hat{n}_i, \hat{y}_i) | \boldsymbol{\theta}] \stackrel{(2)}{=} \prod_{i \in \mathbb{I}} \mathbb{P}[(\hat{n}'_i, \hat{n}''_i, \hat{n}_i) | \boldsymbol{\theta}] \mathbb{P}[\hat{y}_i | \boldsymbol{\theta}] \quad (6)$$

with (1) because conditionally on  $\boldsymbol{\theta}$ , the measures are independent, and (2) because we consider that the measures for immature and mature cells are obtained through two independent experiments. We do not have an analytical

expression for the posterior. We approximate the distribution by sampling from it with a Markov Chain Monte Carlo (MCMC) algorithm, more specifically the Metropolis-Hastings algorithm. It relies on a proposal distribution that we choose as a multivariate normal distribution with zero mean and  $\Sigma$  as covariance matrix. Potentially, the parameter space dimension is large. In that case, the Metropolis-Hastings algorithm often proves inefficient, as it becomes complicated to define a covariance matrix that allows a good convergence of the algorithm. Adaptive algorithms have been proposed to circumvent this problem [14]. We propose here an alternative and simpler method which proved very efficient in practice. We start by learning the covariance matrix of the proposal distribution and by choosing a starting point of the Markov chain located at the maximum a posteriori using the CMA-ES method [12] as detailed below.

### 3.2 Setting the MCMC algorithm using CMA-ES method

The CMA-ES (Covariance Matrix Adaptation - Evolution Strategy) algorithm is a stochastic method for optimization that gives good results in a wide range of problems, including problems that are non-linear, non-separable and in high dimension [12]. This algorithm searches for the maximum of a function over generations. At each generation, a sample of  $\lambda$  individuals (i.e. parameter vectors) is generated, according to a multidimensional normal distribution whose mean and variance-covariance matrix is computed from the selected individuals of the previous generation. Among these  $\lambda$  offsprings, we select  $\mu$  (those which give the highest values for the posterior). These are the ones we use for the next generation. This continues until we reach the maximum number of generations  $n_g$  or if the estimated value for the maximum no longer changes sufficiently over the generations. The basic equation for sampling the individuals, for generation number  $g = 0, 1, 2, \dots, n_g$  reads [13]:

$$\theta_k^{(g+1)} \sim \mathbf{m}^{(g)} + \sigma^{(g)} \mathcal{N}(\mathbf{0}, \mathbf{C}^{(g)}) \quad \text{for } k = 1, \dots, \lambda \quad (7)$$

with  $\mathbf{m}^{(g)}$  the mean value of the search distribution,  $\sigma^{(g)}$  the step-size and  $\mathbf{C}^{(g)}$  the covariance matrix at generation  $g$ . We use this algorithm to find the maximum a posteriori (MAP) of the distribution defined in eq. (5). We use it as starting point for our MCMC algorithm. Moreover, learning the covariance matrix in the CMA-ES is analogous to learning the inverse Hessian matrix in a quasi-Newton method [13]. Thus, we choose for the covariance matrix of the Metropolis proposal the one learned by the CMA-ES method and set  $\Sigma = \mathbf{C}^{(n_g)}$ .

## 4 Results

### 4.1 Parameter estimation

The model presented above in §2.3 potentially has a large number of parameters. However, in this paper, we estimate parameters using data from a single

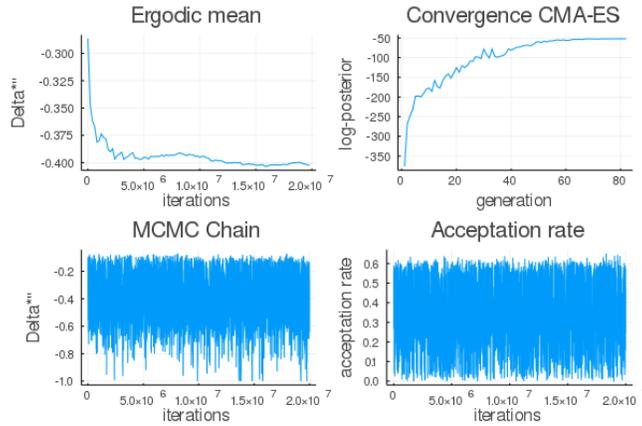
**Table 1.** Parameter estimation

Parameter	Mean	Variance	range	Parameter	Mean	Variance	range
$\gamma$	1.51e-03	8.91e-07	[0.0, 0.05]	$\eta''$	0.123	1.28e-03	[0.0, 2.0]
$\beta$	2.90e-02	1.47e-04	[0.0, 0.05]	$\gamma'_*$	2.43e-02	1.59e-04	[0.0, 0.05]
$\alpha_*$	2.71e-02	1.56e-04	[0.0, 0.05]	$\Delta''_*$	-0.404	2.30e-02	[-1, 1]
$k_{m*}$	0.513	7.51e-02	[0.05, 1.0]	$\eta'$	1.44e-02	3.73e-05	[0, 2]
$k''_m$	11.894	6.48	[1.0, 20.0]	$\Delta'_*$	-0.694	4.54e-2	[-1, 1]

patient. In order to reduce the risk of over-fitting, it is necessary to make several assumptions concerning the parameters, in particular regarding those that vary among the wt, het or hom cell subtypes, with or without IFN $\alpha$  effect. The various simplifying assumptions are as follows:  $\alpha = \alpha' = \alpha'' = 0.00375$  (considering that 150 HSCs are recruited each day, among around 40.000 over 100.000 [15, 11], to contribute to hematopoiesis and that this is independent from the cell type),  $\gamma_* = \gamma$  and  $\beta_* = \beta$  (meaning that the drug does not target healthy HSCs),  $\delta_i = \delta'_i = \delta''_i = \delta_{i*} = \delta'_{i*} = \delta''_{i*} = 1/6$  and  $\delta_m = \delta'_m = \delta''_m = \delta_{m*} = \delta'_{m*} = \delta''_{m*} = 1/3$  (knowing that wild-type granulocytes have a lifespan of the order of magnitude of the day),  $k''_{i*} = k'_{i*} = k_{i*} = 1$  (meaning that IFN $\alpha$  does not influence the proliferation for immature cells),  $k''_i = k'_i = k_i = 1$  (meaning that the mutation does not give a competitive advantage at the progenitor level),  $\chi'' = \chi' = \gamma/(\gamma + \beta)$ ,  $\beta_*' = \beta_*'' = \beta$ ,  $k''_{m*} = k_{m*} = k'_{m*}$  (meaning that IFN $\alpha$  affects equally all mature cells),  $k'_m = k''_m$  and  $\gamma'_* = \gamma''_*$  (meaning that there is only little difference between het and hom cells) and finally  $\alpha'_* = \alpha''_* = \alpha_*$  (meaning that IFN $\alpha$  affects equally the HSC quiescence exit). In the end, we have 10 parameters to estimate. Applying the method described in section 3, and burning the first samples, we obtain the estimations given in Table 1.

## 4.2 Convergence of the algorithms

The validity of the results presented in the previous paragraph depends on the convergence of the algorithms used. Following the method presented in section 3, we start by estimating the maximum a posteriori (MAP) using the CMA-ES algorithm. In Figure 2, top right, we observe the logarithmic value of the MAP over the different generations of the CMA-ES algorithm (where we sample 50 new individuals at each generation). The algorithm converges after 80 generations. Next, we use an MCMC sampling method, where the covariance matrix of the Metropolis proposal is learned using the CMA-ES method and rescaled to obtain a correct acceptance rate, as shown at the bottom right of Figure 2. We iterate over a sufficient number of times until the convergence of the ergodic means of the parameters. In the upper left corner of Figure 2, we see the example of the ergodic mean for  $\Delta''_*$ . At the bottom left of the same figure, we can also observe the chain generated for this parameter, showing a correct exploration of the parameter space.

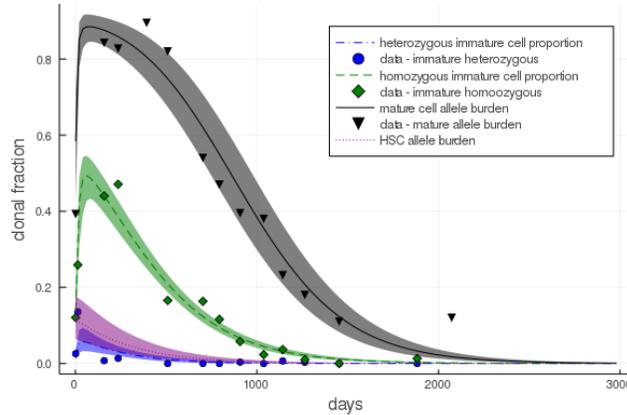


**Fig. 2.** Top left: ergodic mean of the parameter  $\Delta''$  illustrating the convergence of the MCMC schema; bottom left: MCMC chain of the parameter  $\Delta''$  showing how the parameter space is explored; bottom right: variation of the acceptance rate with iterations, it fluctuates around 0.4; top right: log-value of the posterior as a function of the generation with the CMA-ES algorithm

## 5 Discussion

### 5.1 Effect of IFN $\alpha$

As we can see in Figure 3, obtained after propagating uncertainties using a Monte-Carlo method, IFN $\alpha$  affects the hematopoiesis dynamics of the patient in a positive manner: it reduces the number of cancer cells not only in the immature and mature compartments but also in HSCs. We can observe this phenomenon looking at the dotted purple curve which corresponds to the HSC allelic burden defined as follows:  $y_{HSC}(t) = \frac{0.5 \cdot (N_1'(t) + N_2'(t)) + N_1''(t) + N_2''(t)}{N_1(t) + N_1'(t) + N_1''(t) + N_2(t) + N_2'(t) + N_2''(t)}$ . After estimating the model parameters from real data, we can propose some hypotheses on the effect of IFN $\alpha$  on HSCs. First, our model suggests that it may play a role in altering the balance between symmetrical, asymmetrical and differentiated divisions of stem cells and in promoting differentiated divisions. Indeed, almost surely, the values of the parameters  $\Delta''$  and  $\Delta'$  are negative. Their posterior distributions, as we can see on the left of figure 4, are centred around the mean values  $-0.404$  and  $-0.694$  respectively. Secondly, we show that IFN $\alpha$  could promote stem cell proliferation by increasing the value of the parameter  $\alpha$ . We estimate that  $\mathbb{P}[\alpha_* > \alpha] = 0.989$ . Lastly, IFN $\alpha$  seems to promote the quiescence exit of the mutated stem cells, favouring the passage from compartment 1 to 2. We have  $\mathbb{E}[\gamma''/\gamma] = 18.8$  and in Figure 4, on the right, we visualize the distribution of this ratio. Next, we should confirm these results by performing the study on a more significant number of patients. This step could also help understand whether one of these three effects is more important than the two others.



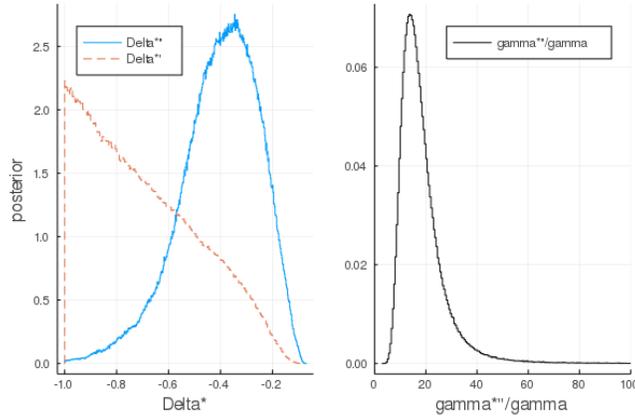
**Fig. 3.** Dynamics of the mutated clonal fraction for the patient under treatment and comparison with real data. The shaded bands represent 95% confidence intervals.

## 5.2 Adding patients

Despite the many simplifications made in paragraph 4.1 to reduce the number of parameters to be estimated, it remains quite large compared to the number of data in our possession, leading to a risk of over-fitting. The results obtained and discussed should, therefore, be confirmed based on data from the entire cohort. This way, we could understand what the dominant effect of  $\text{IFN}\alpha$  on all patients is. However, most of the model parameters are not generic but patient-specific. Multiplying the number of patients will not increase the robustness if it only results in the multiplication of the number of parameters. To circumvent this problem, we will extend the proposed Bayesian framework with a hierarchical model. The hierarchical model hinges on the modification of the expressions for the priors by adding hyper-parameters. With these new hyper-parameters, we can bind the parameters of all patients, which introduces robustness in the parameter estimation by reducing the variance on the results [16].

## 5.3 Pharmacodynamics

The model presented in this article aims to test hypotheses concerning the impact of  $\text{IFN}\alpha$  on hematopoiesis and cancer cells. We built the model on the assumption that the treatment acts on the model parameters in a static way, i.e. the model parameters will potentially take new values as soon as  $t \geq 0$ , date of the beginning of the treatment, and will keep these values throughout the therapy. However, as we mentioned in the paragraph 2.1, the  $\text{IFN}\alpha$  dosage evolves throughout the patient follow-up. The changes in the dosage might explain, for example, the last value observed for the mature cell allele burden on graph 3. The next step is to extend our model so that it can dynamically take into account changes in posology in order to understand how dosage affects the patient's



**Fig. 4.** On the left, posterior distributions of the parameters  $\Delta''$  and  $\Delta'_*$ . On the right, posterior distribution of the ratio  $\gamma''/\gamma$

response to treatment. This could also help to determine optimal posology to achieve molecular remission in patients.

## 6 Conclusion

Based on biological knowledge, we proposed a hematopoiesis model in order to test the effect of the IFN $\alpha$  treatment on mutated hematopoietic cells, primarily stem cells. We designed this compartmental model in coherence with observation data of the experimental protocol of a prospective medical study at Gustave Roussy cancer research institute. Our data come from a patient followed for several years. We then modelled the uncertainties in the experimental measurements and implemented a parameter estimation method based on Bayesian statistics. Due to the complexity of the hematopoiesis model and the measurement noise model, the parameter posterior distribution were numerically approximated. To do this, we developed a method combining a classical MCMC-type algorithm initialized and configured from the results of an optimization algorithm, the CMA-ES method. Our method eased the convergence of the Metropolis-Hastings algorithm, especially in high dimension. We were then able to test our biological hypotheses and see to what extent IFN $\alpha$  allowed the molecular remission of the patient by acting on the stem cell stock. Our preliminary results seem to confirm our hypothesis that IFN $\alpha$  acts on stem cells by promoting the quiescence exit of stem cells, particularly mutated stem cells, and by leading to the exhaustion of the stock of mutated stem cells by promoting the differentiated division mechanism. These results will be confirmed by studying all patients in the cohort and by the implementation of a hierarchical Bayesian framework which enables to have more robust results. Once confirmed the impact of IFN $\alpha$  on stem cells, the

next step will be to study the effect of posology in order to provide patients with personalized treatments.

## References

1. Vainchenker, W., Kralovics, R. (2017). Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. *Blood, The Journal of the American Society of Hematology*, 129(6), 667-679.
2. Yacoub, A., Mascarenhas, J., Kosiorek, H., Prchal, J. T., Berenzon, D., ... & Finazzi, M. C. (2019). Pegylated interferon alfa-2a for polycythemia vera or essential thrombocythemia resistant or intolerant to hydroxyurea. *blood*, 134(18), 1498-1509.
3. Gisslinger, H., Klade, C., Georgiev, P., Krochmalczyk, D., ... & Sivcheva, L. (2020). Ropoginterferon alfa-2b versus standard therapy for polycythaemia vera (PROUD-PV and CONTINUATION-PV): a randomised, non-inferiority, phase 3 trial and its extension study. *The Lancet Haematology*, 7(3), e196-e208.
4. Adimy, M., Chekroun, A., Touaoula, T. M. (2015). Age-structured and delay differential-difference model of hematopoietic stem cell dynamics.
5. Xu, J., Koelle, S., Guttorp, P., Wu, C., Dunbar, C., Abkowitz, J. L., Minin, V. N. (2019). Statistical inference for partially observed branching processes with application to cell lineage tracking of in vivo hematopoiesis. *The Annals of Applied Statistics*, 13(4), 2091-2119.
6. Kimmel, M. (2014). Stochasticity and determinism in models of hematopoiesis. In *A systems biology approach to blood* (pp. 119-152). Springer, New York, NY.
7. Marciniak-Czochra, A., Stiehl, T., Ho, A. D., Jäger, W., Wagner, W. (2009). Modeling of asymmetric cell division in hematopoietic stem cells—regulation of self-renewal is essential for efficient repopulation. *Stem cells and development*, 18(3), 377-386.
8. Colijn, C., Mackey, M. C. (2005). A mathematical model of hematopoiesis—I. Periodic chronic myelogenous leukemia. *Journal of Theoretical Biology*, 237(2), 117-132.
9. Michor, F., Hughes, T. P., Iwasa, Y., Branford, S., Shah, N. P., Sawyers, C. L., Nowak, M. A. (2005). Dynamics of chronic myeloid leukaemia. *Nature*, 435(7046), 1267-1270.
10. Catlin, S. N., Abkowitz, J. L., Guttorp, P. (2001). Statistical Inference in a Two-Compartment Model for Hematopoiesis. *Biometrics*, 57(2), 546-553.
11. van der Wath, R. C., Wilson, A., Laurenti, E., Trumpp, A., Lio, P. (2009). Estimating dormant and active hematopoietic stem cell kinetics through extensive modeling of bromodeoxyuridine label-retaining cell dynamics. *PLoS one*, 4(9), e6972.
12. Hansen, N. (2006). The CMA evolution strategy: a comparing review. In *Towards a new evolutionary computation* (pp. 75-102). Springer, Berlin, Heidelberg.
13. Hansen, N. (2016). The CMA evolution strategy: A tutorial. arXiv preprint arXiv:1604.00772.
14. Andrieu, C., & Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and computing*, 18(4), 343-373.
15. Lee-Six, H., Øbro, N. F., Shepherd, M. S., Grossmann, S., Dawson, K., Belmonte, M., ... & O'Neill, L. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature*, 561(7724), 473-478.
16. Llamosi, A., Gonzalez-Vargas, A. M., Versari, C., Cinquemani, E., Ferrari-Trecate, G., Hersen, P., & Batt, G. (2016). What population reveals about individual cell identity: single-cell parameter estimation of models of gene expression in yeast. *PLoS computational biology*, 12(2), e1004706.