

RELATÓRIO TÉCNICO

Universidade Federal Rural da Amazônia

Belém, Pará, 23/05/2020

Redes Neurais Artificiais e Modelagem Matemática nas Previsões Epidemiológicas para os Casos de Infecção por COVID-19

Uma Ferramenta de Apoio à Tomada de Decisão em Políticas de Saúde Pública no Estado do Pará

APRESENTAÇÃO

O coronavírus SARS-CoV-2, anteriormente denominado de 2019-nCoV, é o patógeno causador da Síndrome Respiratória Aguda Grave do Coronavírus 2, amplamente difundida como COVID-19 e possui uma disseminação extremamente rápida, forçando a aplicação de novas políticas públicas para seu enfrentamento. Diversas técnicas de predição de casos futuros de contágio e óbitos já foram apresentadas em diferentes regiões do planeta. Esta pesquisa é coordenada pela Pró-Reitoria de Extensão da Universidade Federal Rural da Amazônia – UFRA, e tem como objetivo propor modelos de previsão de curto prazo baseado em Análise dos Componentes Principais e no Aprendizado de Máquina por Redes Neurais Artificiais, capazes de estimar o número de casos de contágio e de óbitos causados pelo SARS-CoV-2. Os modelos propostos são baseados em dados, onde toda inferência é feita a partir do conhecimento descoberto a priori. Ademais, este estudo também apresenta a expectativa do dimensionamento dos recursos físicos e humanos em instalações hospitalares. O estudo apresenta resultados para todo o estado, bem como microrregiões do Pará, por meio da análise de séries temporais que servem de suporte à tomada de decisão das autoridades públicas.

HISTÓRICO CIENTÍFICO DA PESQUISA

O primeiro estudo realizado utilizou a técnica de Redes Neurais Artificiais (RNAs) para fazer previsão de curto prazo dos casos de contágio e letalidades para um período de 30 dias, a contar de 18/03/2020, data da primeira ocorrência de COVID-19 no estado. Os resultados previstos foram publicados no International Journal of Development Research (vol. 10, issue 04, 35416-35421, April, 2020, D.O.I. 10.37118/ijdr.18717.04.2020). A Tabela 1 resume os resultados destas previsões em contraste com os dados reais observados, bem como a taxa de acerto.

Tabela 1. Primeiras Previsões do Modelo Neural.

Ocorrências	Previsto	Observado	% de Acerto	Data
Nº de contágios	8.895	9.059	98,19	12/maio
Nº de óbitos	500	494	98,79	05/maio

No segundo momento, visando uma maior acurácia das previsões, o modelo foi reajustado para estimativas de 7 dias. Os resultados estão colecionados em um Projeto de Extensão Universitária intitulado “Informes sobre o Comportamento da COVID-19 utilizando Modelagem Matemática e Inteligência Computacional no estado do Pará”, cadastrado na Pró-Reitoria de Extensão da UFRA, sob o Código PD002-2020. Dois boletins já foram publicados até o momento: BOLETIM COVID PA #1 (D.O.I. 10.13140/RG.2.2.28930.09924) e BOLETIM COVID PA #2 (D.O.I. 10.13140/RG.2.2.19019.41768) e estão disponíveis em <<https://proex.ufra.edu.br/boletimcovidufra>>.

A partir do momento em que a SESPÁ passou a reportar os dados de contágio e óbitos por COVID-19 separando os casos diários dos casos acumulados, foi possível melhorar o modelo inteligente de estimativa. Desta forma, as previsões passaram a ter um grau de certeza mais elevado. O BOLETIM COVID PA #2 já reflete esta mudança na forma de divulgação dos dados epidemiológicos. Os boletins foram gerados com dados acumulados a partir do dia 18/03/2020.

METODOLOGIA

Dados

Os dados utilizados neste estudo são públicos, fornecidos pela Secretaria de Estado de Saúde do Pará (<http://www.saude.pa.gov.br/coronavirus/>), pelo Ministério da Saúde (<https://saude.gov.br/>) e pelos relatórios da Organização Mundial da Saúde (<https://www.who.int/docs/>). Foram identificados os padrões comportamentais da sociedade, extraídos para análise os casos ocorridos no estado do Pará e comparados com os estratos nacionais, continentais e mundiais. Com base neste banco de dados criado, foi usada a série histórica com informações diárias das ocorrências a partir de 18/03/2020, data do primeiro caso de COVID-19 confirmado no Pará, até o dia 22/05/2020.

Redes Neurais Artificiais (RNAs)

As RNAs são capazes de realizar a aproximação de funções contínuas e podem ser usadas com sucesso na previsão de séries temporais com a vantagem de aproximar qualquer função não linear sem nenhuma informação prévia sobre as propriedades da série de dados. Como as Redes Neurais podem aproximar funções não lineares com um conjunto limitado de parâmetros, elas servem como uma ferramenta eficaz para aproximar e estimar os efeitos de uma quarentena em combinação com os modelos epidemiológicos analíticos.

Simulações

1. Óbitos e Casos Confirmados por Redes Neurais Artificiais (RNA).

Para se determinar as variáveis que seriam utilizadas na composição dos neurônios artificiais da camada de entrada das redes neurais, foi aplicado o método de Análise de Componentes Principais (ACP), que processou o conjunto de 153.135 observações. Dias Após o Início da Contaminação: numérico inteiro; Casos Confirmados: numérico inteiro; Óbitos: numérico

inteiro. As variáveis consideradas para o treinamento e validação das RNAs, considerando-se as entradas e saídas dos ciclos de treinamento, são qualitativas, quantitativas e discretas. As variáveis que compoem os neurônios da camada de saída das RNAs são os Casos Confirmados (CC) e os Óbitos (OB) por COVID-19.

2. *Estimativa de Recursos Hospitalares por Redes Neurais Artificiais (RNA).*

Foram treinadas RNAs para inferir, a partir da quantidade diária de óbitos estimadas na etapa anterior; das demandas hospitalares por médico, enfermeiro, fisioterapeutas e técnicos de enfermagem; a partir de dados técnicos da OMS e científicos de Nova York (EUA) e Wuhan (China), considerando a taxa de uso de leitos no pico pandêmico de 2,5 leitos de UTI/10.000 adultos, e a proporção da população por faixa etária do estado do Pará.

As medidas de precisão do ajustamento das redes neurais artificiais treinadas e validadas para cada etapa da pesquisa são apresentadas na Tabela 2.

Tabela 2. Medidas de precisão do ajustamento das RNAs utilizadas na pesquisa.

Características das RNAs Selecionadas		Mundo	Estado do Pará	Região Metropolitana de Belém	Dimensionamento de Recursos Hospitalares
Arquitetura		MLP 8-12-2	MLP 28-10-2	MLP 1-8-4	MLP 1-4-10
Treinamento	Correlação	0,98064	0,99866	0,99174	0,99923
	Erro	0,00150	0,00002	0,00233	0,00218
Validação	Correlação	0,98185	0,99702	0,98054	0,99923
	Erro	0,00175	0,00007	0,00793	0,00218
Algoritmo de Treinamento		BFGS 3927	BFGS 10000	BFGS 430	BFGS 37
Função de Ativação	Camada Oculta	Logistic	Exponential	Logistic	Tanh
	Camada de Saída	Identity	Logistic	Exponential	Exponential

RESULTADOS PRELIMINARES

Com o objetivo de identificar os principais atributos a serem utilizados na RNA, os autovetores da análise de componentes principais, indicaram que a expressão das variáveis deveria ser separada em dois grupos, mundial e regiões das Américas, com ligação de extração comportamental entre eles (Figura 1).

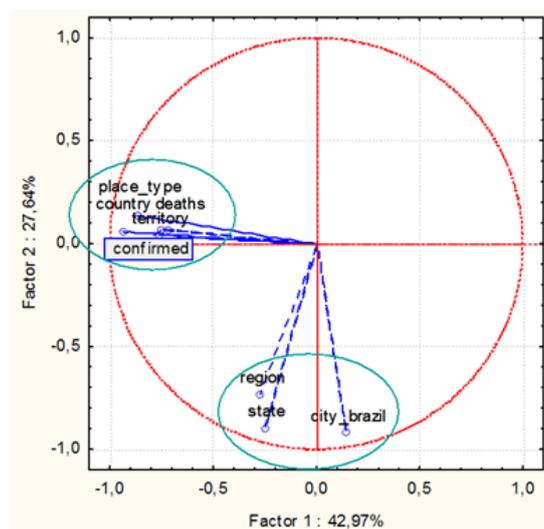


Figura 1. Círculo de correlação com agrupamento multivariado nos dois componentes principais com maior responsabilidade na variação dos dados.

A Figura 2 mostra que os números de contágios e óbitos previstos estão em consonância com o observado, o que demonstra a acurácia do modelo. Estes resultados estão reportados nos dois boletins publicados a partir dos dados acessados no site da SESPA.

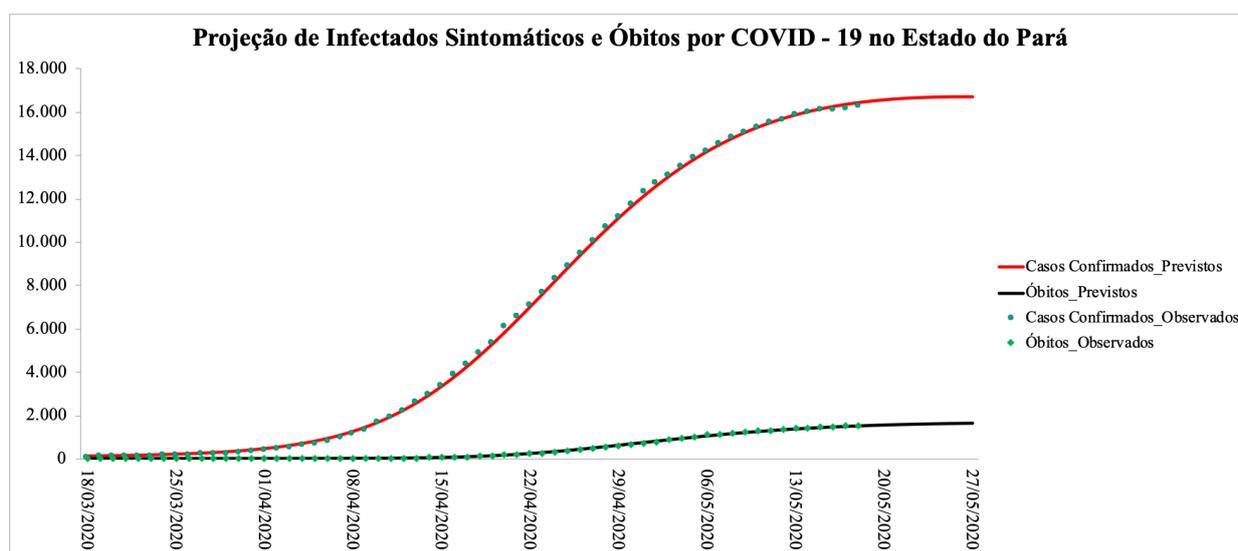


Figura 2. Observação dos casos confirmados com a doença COVID-19 e os óbitos causados por esta em comparação a previsão realizada, no Estado do Pará.

É importante destacar que, ao se usar métodos baseados em dados, as previsões podem ser afetadas por variáveis externas, como se observa na quantidade de óbitos na Tabela 3, que não alcançou o número de óbitos preditos pelo modelo, provavelmente, por influência de fatores como o comportamento da população e sua adesão ao isolamento social, bem como de ações tomadas pela administração pública, como a transformação dos serviços na policlínica para o atendimento exclusivo para COVID-19, abertura do Hospital Regional Abelardo Santos para pronto atendimento de porta aberta, nos quais além do

atendimento médico foram também disponibilizadas medicações para o tratamento da doença em situações que não requeressem internação.

Tabela 3. Quantificação dos casos/óbitos previstos e casos/óbitos confirmados no período de 13 a 20 de maio de 2020. Destacando-se a ocorrência de -11,5% (205) de óbitos. Estado do Pará

Previsão de Casos Confirmados de 13 a 20 de maio.

Data	Previsão de Casos Confirmados			Casos Confirmados Publicados	Diferença Mediana	
	Otimista	Mediano	Pessimista		Casos	%
13-mai	9.119	10.031	11.034	10.344	-313	-3,0%
14-mai	9.932	10.926	12.018	11.479	-553	-4,8%
15-mai	10.811	11.892	13.081	12.626	-734	-5,8%
16-mai	11.759	12.935	14.229	13.464	-529	-3,9%
17-mai	12.786	14.065	15.472	14.201	-136	-1,0%
18-mai	13.897	15.287	16.816	15.467	-180	-1,2%
19-mai	15.099	16.609	18.269	17.177	-568	-3,3%
20-mai	16.398	18.038	19.841	18.929	-891	-4,7%

Previsão de Óbitos de 13 a 20 de maio.

Data	Previsão de Óbitos			Óbitos Publicados	Diferença Mediana	
	Otimista	Mediano	Pessimista		Óbitos	%
13-mai	882	970	1.067	1.022	-52	-5,1%
14-mai	980	1.078	1.185	1.095	-17	-1,6%
15-mai	1.087	1.196	1.316	1.175	21	1,8%
16-mai	1.205	1.326	1.458	1.223	103	8,4%
17-mai	1.335	1.468	1.615	1.280	188	14,7%
18-mai	1.477	1.625	1.787	1.392	233	16,7%
19-mai	1.633	1.796	1.976	1.554	242	15,6%
20-mai	1.803	1.983	2.182	1.778	205	11,5%

Ao comparar a expectativa de demanda por leitos de UTI com a lista de espera por leitos na região metropolitana de Belém, bem como com a atual ocupação de leitos de UTI na rede estadual de saúde pública (84,56% - 330 leitos ocupados), percebe-se que os números para o atual momento são condizentes com a realidade e com uma tendência decrescente, em razão da redução da curva de óbitos na região metropolitana de Belém (Tabela 4).

Tabela 4. Projeções e Demanda de Recursos Hospitalares, no Estado do Pará..

Recurso	06/maio/2020	13/maio/2020	20/maio/2020	27/maio/2020
UTI	1.196	957	385	252
Leitos Clínicos	3.707	2.865	1.258	761
Médico	655	518	215	138
Fisioterapeuta	180	144	57	38
Enfermeiro	721	570	236	151
Técnico de Enfermagem	1.561	1.192	532	330

Com os dados disponibilizados diretamente pela Secretaria de Saúde do Estado, avaliou-se o modelo proposto, e observou-se que ele (expresso com a linha vermelha) se aproximou da quantidade de óbitos registrada (Figura 3). Tanto o modelo quanto os dados disponibilizados demonstram que há uma redução significativa no número de óbitos, considerando-se ainda, que o período de maior risco de óbitos ocorreu entre os dias 28 de abril e 01 de maio.

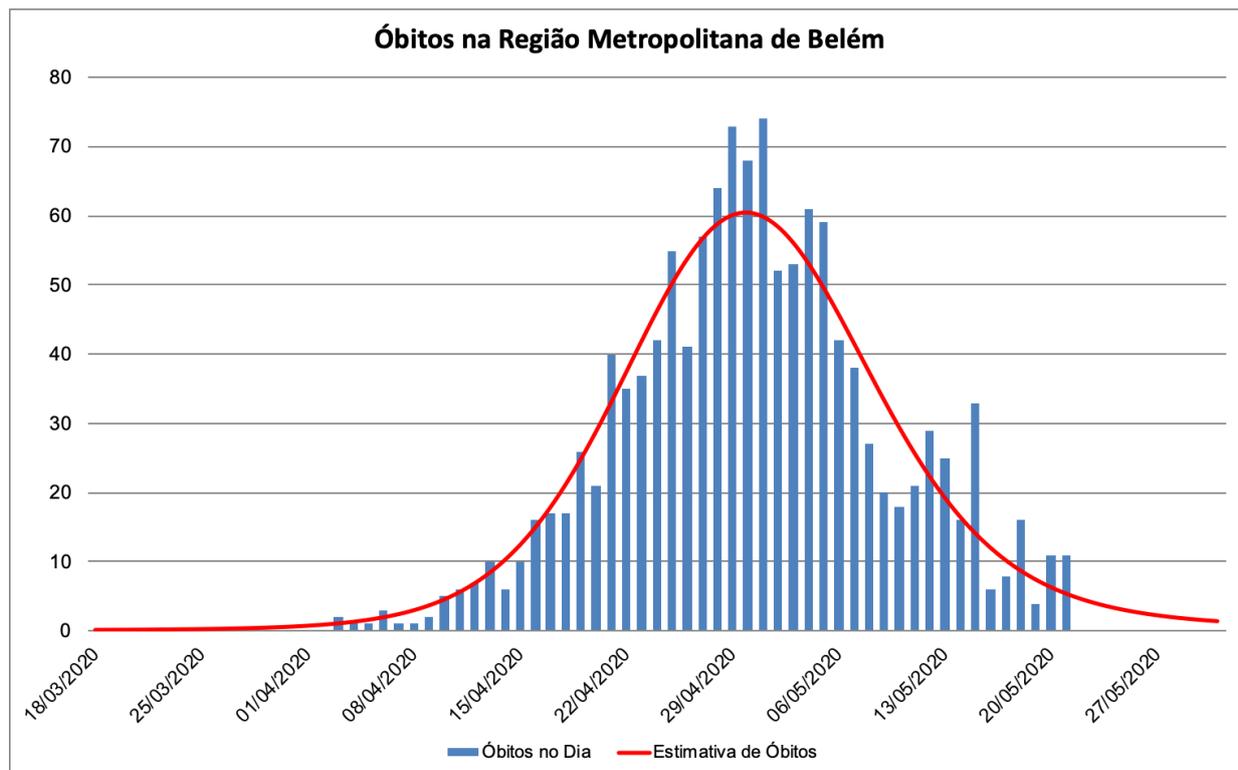


Figura 3. Comparação dos casos registrados na SESPA (barras azuis) com o previsto no modelo (linha vermelha) quanto a quantidade de óbitos na Região Metropolitana de Belém.

O modelo também demonstrou acurácia na previsão dos casos confirmados, onde se observa uma redução mais lenta, conforme prevista pelo modelo (linha vermelha) na Figura 4. O comportamento da curva de predição é assimétrica e, portanto, a redução da evolução diária de casos é mais suave e gradual, quando comparada com a evolução da mesma até o pico pandêmico. Pode-se ainda verificar que o período de maior risco de contágio ocorreu entre os dias 16 a 20 de abril.

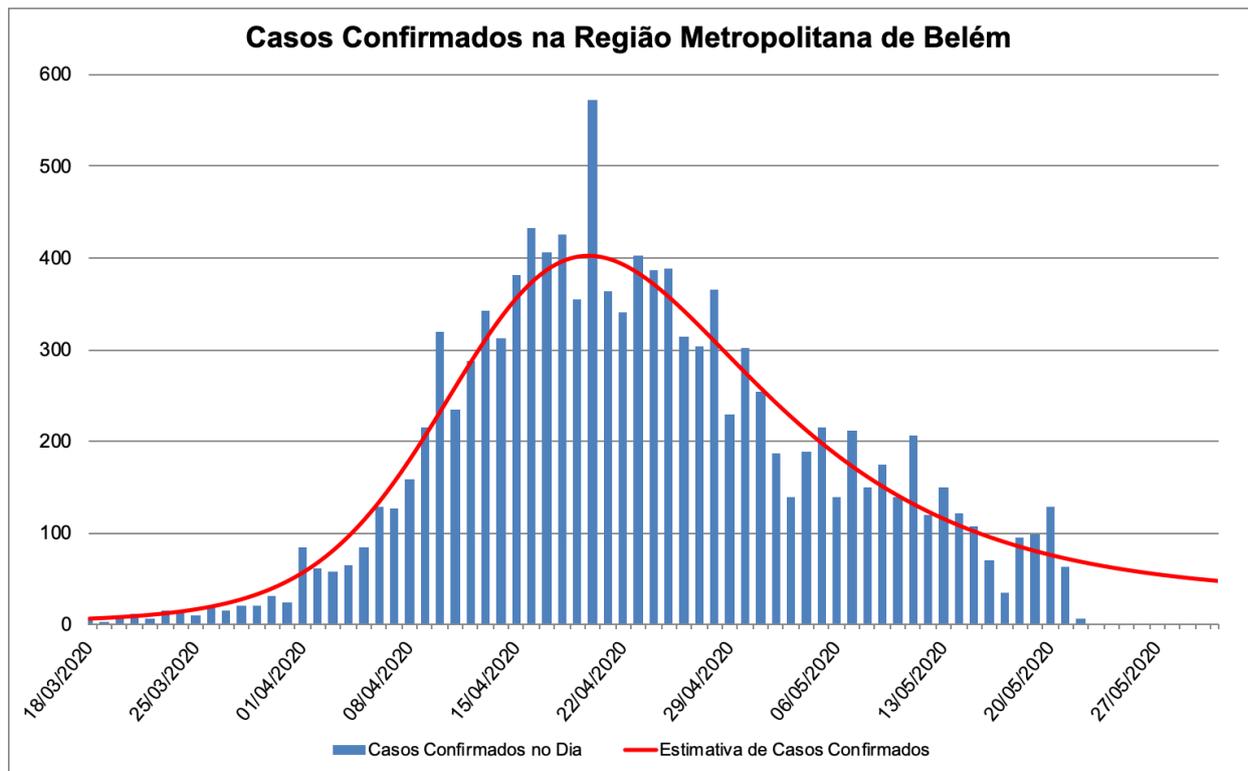


Figura 4. Comparação dos casos registrados na SESPA (barras azuis) com o previsto no modelo (linha vermelha) quanto a quantidade de casos confirmados, na Região Metropolitana de Belém.

Ainda para casos confirmados, ao se avaliar os dados acumulados desde o início dos registros, na Figura 5, pode-se visualizar a formação de um platô nas curvas real e prevista, indicando uma estabilização de casos confirmados com síndrome respiratória aguda grave na região metropolitana de Belém, com acréscimos suaves de novos casos.

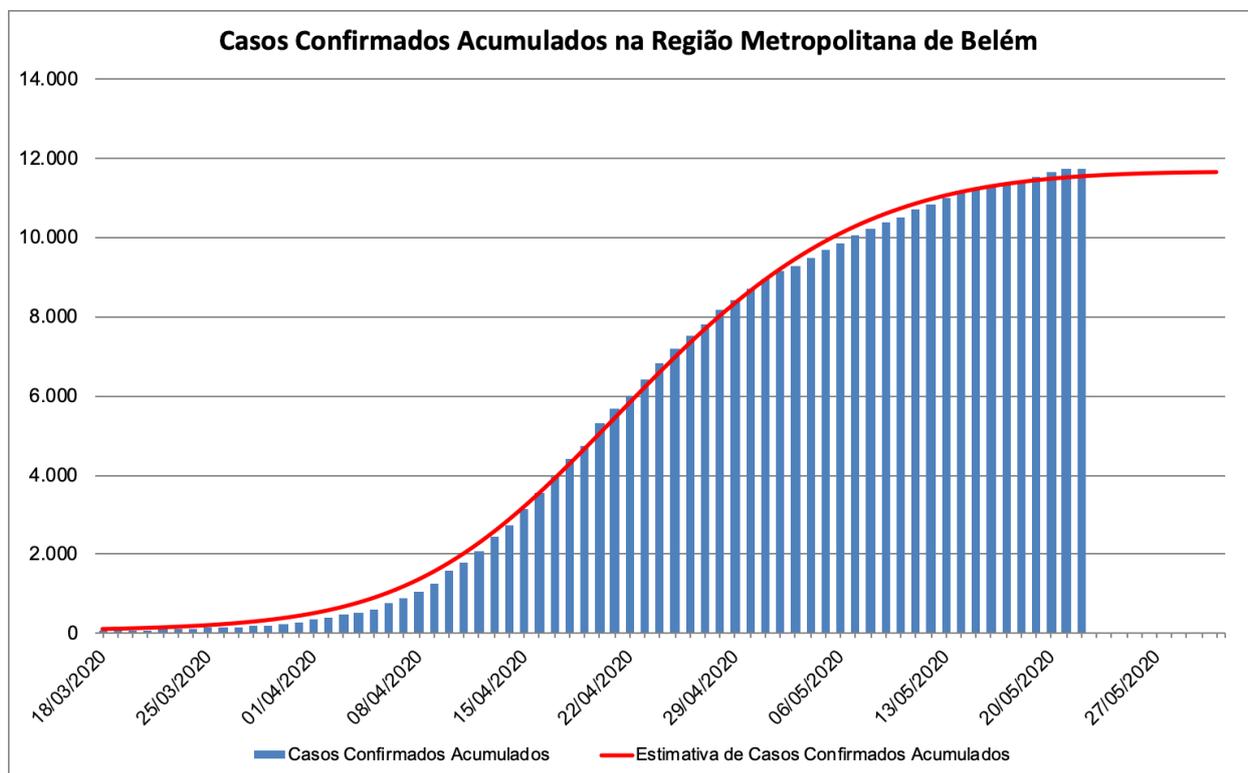


Figura 5. Comparação dos casos registrados na SESP (barras azuis) com o previsto no modelo (linha vermelha) quanto a quantidade de casos confirmados acumulados, na Região Metropolitana de Belém.

CONSIDERAÇÕES FINAIS

As previsões mostraram, entre os dias 16 a 20 de abril, uma tendência de aumento expressivo do número de casos confirmados de COVID-19, bem como no intervalo entre os dias 21 a 26 de abril, a tendência de redução de contágio.

É importante ressaltar que outras variáveis não previstas podem influenciar nas projeções e no tempo de duração da pandemia. Outro aspecto que pode afetar as previsões é o avanço da pandemia em direção aos municípios do interior do Estado do Pará e a taxa de adesão da população às medidas de combate ao coronavírus, como o isolamento social.

As subnotificações de casos confirmados e óbitos que se somam a datas passadas podem invalidar os cenários atuais de previsões e, portanto, é necessário o reprocessamento recorrente de dados com a nova matriz de realidade.

A região metropolitana de Belém apresenta uma tendência de redução na contaminação e óbitos por COVID-19, bem como na sua demanda por recursos hospitalares. Este fato, na atual conjuntura, permite afirmar que o dimensionamento destes recursos está condizente com a capacidade de suprimento do estado.

EQUIPE DE TRABALHO

A equipe que desenvolve esta pesquisa é multidisciplinar e conta com profissionais das áreas da Saúde, Engenharia, Biologia Computacional, Modelagem Matemática e Inteligência Computacional.

A equipe de trabalho que assina esta nota técnica é composta por membros do Grupo de Pesquisa (CNPq/UFRA) Núcleo de Pesquisa em Computação Aplicada.

Prof. Carlos Maneschy
Secretário - SECTET/PA
Presidente Fapespa
c.maneschy15@gmail.com

Prof. Jonas Castro
Pró-Reitor Adjunto de Extensão
UFRA
jonas.castro@ufra.edu.br

Prof. Marcel Botelho
Reitor
UFRA
marcel.botelho@ufra.edu.br

Prof. Marcus Braga
Campus de Paragominas
UFRA
marcus.braga@ufra.edu.br

Prof. Rommel Ramos
Instituto de Ciência Biológicas
UFPA
rommel.ramos@ufpa.br