Instituto de MATEMÁTICA E ESTATÍSTICA UFRGS

DEPARTAMENTO DE ESTATÍSTICA
E-mail: dest@mat.ufrgs.br

Trabalho de Conclusão de Curso

Modelagem de zonas de influência no futebol usando *tracking data*

Luiz Almir Zanella de Paula

2024

Luiz Almir Zanella de Paula

Modelagem de zonas de influência no futebol usando *tracking data*

Trabalho de Conclusão apresentado à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientadora: Prof[a]. Dr[a]. Márcia Helena Barbian

Porto Alegre
Fevereiro de 2024

Luiz Almir Zanella de Paula

Modelagem de zonas de influência no futebol usando *tracking data*

Este Trabalho foi julgado adequado para obtenção dos créditos da disciplina Trabalho de Conclusão de Curso em Estatística e aprovado em sua forma final pelo(a) Orientador(a) e pela Banca Examinadora.

Orientadora:_____
Prof[a]. Dr[a]. Márcia Helena Barbian, UFRGS
Universidade Federal do Rio Grande do Sul, Porto Alegre, RS

Banca Examinadora:

Prof. Dr. Marcio Valk, UFRGS
Doutor pela Universidade Estadual de Campinas, Campinas, SP

Prof. Dr. Rodrigo Citton Padilha dos Reis, UFRGS
Doutor pela Universidade Federal de Minas Gerais, Belo Horizonte, MG

Porto Alegre
Fevereiro de 2024

"You miss 100% of the shots you don't take
- Wayne Gretzky"
- Michael Scott
""All models are wrong, but some are useful"
- George Box"

Agradecimentos

Agradeço primeiramente ao meu pai e a minha mãe pelo carinho e dedicação que vocês me deram a vida toda. Ainda mais por terem me apoiado quando este maluco que vos fala decidiu abandonar a carreira de advogado para se aventurar e estudar algo que realmente gostava.

À minha avó, Neuza, que infelizmente não está aqui para me ver nesse momento mas que me dava bença todo dia e dizia que rezava para o Padre Reus me arrumar um emprego. Eu sinto a tua falta, vó.

Aos meus irmãos, César e Ana, por serem parceiros e sempre sermos suporte uns dos outros.

Aos meus tios, Carla e Júlio, que me aturam quando invado a casa de vocês para pegar água com gás ou passar o dia inteiro perguntando se vocês tem algum doce para me dar. Quem foi Buda perto do que vocês aguentam.

À minha prima Ana Júlia, por ela sim me dar doce sempre que eu peço.

À minha tia, Maria, que sempre fica doida quando assusto ela (o que acontece com uma frequência não muito saudável).

Aos nossos gatos, Caetano, Bethânia e Mabel por me deixarem afofar eles o dia todo e serem os amores da minha vida. Menção honrosa para a Bebê, a Capeta, a Tuts e a Jagua.

À minha querida professora/chefe/orientadora e por fim madrinha, professora Márcia Helena Barbian, por ter me aturado como aluno/bolsista/orientando e agora também afilhado. Lembro que no início do curso eu comentava que queria fazer TCC sobre esportes e me diziam que tu era a pessoa indicada para isso mas eu morria de vergonha de ir falar contigo. Obrigado por todo ensinamento, toda ajuda e paciência ao longo do curso.

Ao professor Márcio Valk, por ser nosso paraninfo, por ter aceitado fazer parte desta banca, além de ter sido um excelente professor de séries temporais e parceiro de projetos extraclasse. Te admiro muito.

Ao professor Rodrigo Citton Padilha dos Reis, por também ter aceitado fazer da banca e por toda a paciência e ensinamentos nas cadeiras de amostragem e dados correlacionados.

Ao professor Cleiton Taufemback e a professora Taiane Prass, por tudo que aprendi com vocês como aluno e como bolsista. Saudades das nossas reuniões.

Aos demais professores do departamento pelo ótimo ensino e contribuição na minha formação. Serei eternamente grato por todos os ensinamentos.

Aos meus colegas formandos, em especial o Igor e a Martha. Obrigado por ouvirem todas as minhas reclamações, e olha que eu gosto de reclamar.

Aos ex-colegas e em breve colegas de profissão, Alisson, Andressa, Gabriel, Ju-

liana, Raquel e Tainá. Obrigado pelas ajudas e parceria ao longo do curso. Vocês são especiais.

Aos amigos da matemática que fiz ao longo do curso, André, Bia, Júlia, Lucas, Luiza e Renatinha. Obrigado pela amizade. Amo vocês.

Aos meus amigos da época do direito (e chegados), Anna, Gabi, Gus, Paloma, Pati e Ruiva. O nosso grupo é o maior motivo de risadas de cada dia.

Aos meus amigos da vida, Cami, Chapa, Chico, Binho e Zeca, obrigado por todos os rolês e pela parceria ao longo dos anos.

Com certeza esqueci de muita gente, mas o que importa é a intenção. Obrigado a todos que de qualquer forma de auxiliaram a chegar até aqui.

Por fim, obrigado à mãe UFRGS pela oportunidade de estudar numa instituição dessa qualidade.

Resumo

Este trabalho tem como objetivo aplicar técnicas de geoestatística em dados de rastreamento de futebol e assim identificar zonas de influência que cada atleta exerce através das ações no momento que estão com a posse de bola, de forma a fornecer uma nova ferramenta para auxiliar uma comissão técnica a identificar jogadores que possam demandar mais atenção num contexto tático. Para isso, foi utilizada metodologia que utiliza um modelo geoestatístico bayesiano espaço-temporal através de técnica desenvolvida por Rue et al. (2009), e utiliza como modelo o *Stochastic Partial Differential Equations* (SPDE). Através deste método, computamos a moda *a posteriori* por meio do *Integrated Nested Laplace Approximation* (INLA). Pela excessiva demanda computacional, o modelo é aplicado a apenas uma partida de futebol. Os resultados obtidos confirmam a existência de dependência espacial e temporal entre os jogadores ao longo de uma partida.

Palavras-Chave: Geoestatística, SPDE, R-INLA, Futebol, Estatística Espacial, Espaço-Tempo.

Abstract

The aim of this work is to apply geostatistical techniques to soccer tracking data and thus identify zones of influence that each player exerts through their actions when in possession of the ball, in order to provide a new tool to help a coaching staff identify players who may require more attention in a tactical context. To do this, we used a methodology that employs a Bayesian spatio-temporal geostatistical model through a technique developed by Rue et al. (2009), and uses the Stochastic Partial Differential Equations (SPDE) as a model. Using this method, we compute the *posteriori* mode using the Integrated Nested Laplace Approximation (INLA). Due to the excessive computational demands, the model is applied to just one soccer match. The results obtained confirm the existence of spatial and temporal dependence between players over the course of a match.

Keywords: Geoestatistical data, SPDE, R-INLA, Soccer, Spacial Statistics, Space-Time.

Sumário

Lista de Figuras

Lista de Tabelas

1 Introdução

1.1 Considerações Iniciais

O futebol é, inquestionavelmente, o esporte mais praticado e acompanhado do mundo, atravessando fronteiras e conectando diferentes culturas. Com uma história que se estende por séculos e uma presença que abrange todos os cantos do planeta. Milhões de adeptos em todo o mundo vivem intensamente as emoções proporcionadas pelas partidas.

De acordo com dados fornecidos pela *Fédération Internationale de Football Association*, a FIFA, entidade máxima do futebol, a Copa do Mundo do Catar em 2022 contou com engajamento de 5 bilhões de pessoas, sendo que a final do torneio, foi acompanhada ao vivo por 1,5 bilhões de pessoas, tornando este um dos eventos esportivos mais assistidos da história (Fédération Internationale de Football Association - FIFA (2023)).

A imprevisibilidade do futebol é uma das características mais intrigantes e cativantes do esporte. Apesar de sua estrutura aparentemente simples, o jogo é altamente dinâmico e suscetível a reviravoltas dramáticas a qualquer momento. A imprevisibilidade não apenas mantém os espectadores em suspense, mas também desafia os jogadores e treinadores a se adaptarem rapidamente às mudanças nas circunstâncias do jogo. Essa natureza imprevisível do futebol é o que torna cada partida única e emocionante, gerando momentos memoráveis e imprevisíveis que alimentam a paixão e o interesse dos fãs em todo o mundo.

1.2 Justificativa

Atualmente, com a popularização do futebol, o nível profissional dos envolvidos nos clubes, como treinadores, departamentos médicos e, inclusive, os próprios jogadores, está em constante ascensão. Nesse sentido, a busca incessante por aprimorar as condições físicas dos atletas e otimizar o posicionamento tático é a motivação por trás do contínuo investimento na análise de dados e na busca de padrões para os jogadores de futebol (Di Salvo et al. (2006)).

As cifras envolvendo o esporte aumentam ano a ano. De acordo com a agência de notícias Reuteurs, na janela de transferências do verão europeu, juntos os times da Premier League inglesa gastaram um total de 2,36 bilhões de libras em transferências de atletas. A liga sozinha foi responsável por 48% dos valores gastos em transações na janela (Reuters (2024)).

A análise estatística no futebol emergiu como uma ferramenta indispensável, proporcionando *insights* valiosos para equipes, treinadores e entusiastas do esporte. A coleta e interpretação de dados estatísticos permitem uma compreensão mais aprofundada do desempenho individual e coletivo em campo, indo além da observação subjetiva. Métricas como posse de bola, passes precisos, chutes a gol, entre outras, são meticulosamente analisadas para avaliar a eficácia tática, identificar padrões de jogo e tomar decisões estratégicas informadas.

A ascensão da análise estatística no futebol não apenas influencia as decisões dentro do campo, mas também redefine a maneira como os torcedores e profissionais do esporte percebem e apreciam a complexidade do jogo, oferecendo uma visão mais precisa e abrangente do que ocorre nas quatro linhas. Dado o grande campo, os numerosos jogadores, a rotatividade limitada de jogadores e a pontuação escassa, o futebol é, sem dúvida, o mais desafiador de analisar de todos os principais esportes coletivos (Liu et al. (2020)).

Inicialmente, a análise estatística de dados de futebol se limitavam a anotações de eventos ocorridos durante um jogo, como: número de gols, número de finalizações, número de passes, etc, os chamados *event data*. A evolução desse tipo de análise surgiu inspirada principalmente por outros esportes, como beisebol, futebol americano e basquete (Fernández (2021)). Além disso, a revolução na capacidade computacional e de armazenamento de dados tornou possível a utilização de modelos mais complexos na análise de dados de futebol, capacitando a melhora na identificação de jogadores com maior qualidade técnica, a análise da dinâmica de uma partida e previsão de resultados dos jogos(Spearman (2018)).

Alguns desses modelos visavam verificar se haveria vantagem em uma equipe atuar como visitante e quais fatores seriam relevantes para se determinar isto (Pollard (2008)). Outros buscam criar modelos preditivos para os mais variados fins e usando diversas técnicas. Um desses trabalhos foi o realizado por (Tsokos et al. (2018)) onde foram comparados modelos usando diversas extensões de Bradley-Terry treinados usando técnicas de *machine learning* com um modelo log-linear hierárquico de Poisson cujos hiperparâmetros foram ajustado usando *Integrated nested Laplace approximations*, o INLA.

Apesar dessa evolução, majoritariamente, os modelos se limitam a entender o jogo através de ações que sejam quantificadas através da posse de bola ou o resultado da partida, como o desenvolvimento de técnicas de previsão de resultado e análise de desempenho de jogadores (Spearman (2018)).

Com o desenvolvimento tecnológico a informação não só dos passes, chutes ou roubos de bola são armazenados, outro tipo de dado foi agregado ao estudo do esporte. Tais informações envolvem o posicionamento de todos os atletas durante a partida, o chamado *tracking data*. Por ser uma tecnologia recente e de alto custo de implementação, ainda é restrita à clubes da elite do futebol mundial, como o Barcelona (Fernández (2021)).

1.3 Objetivos

A proposta da presente pesquisa é avançar além do uso convencional de *event data*, que abrange informações como passes, finalizações e posse de bola, para analisar o comportamento dos demais atletas em esportes coletivos, como o futebol. Pois, no futebol, a relevância não se limita apenas aos jogadores que estão em posse

da bola, uma vez que o comportamento dos demais jogadores, mesmo sem a posse, pode impactar consideravelmente no desenrolar do jogo.

Nesse sentido, é crucial explorar as áreas específicas do campo onde um jogador, mesmo sem a bola, possui maior probabilidade de influenciar o jogo, levando em consideração sua posição estratégica em relação à bola, aos companheiros de equipe e aos adversários. Tais áreas são denominadas como "zonas de influência". Por exemplo, se um jogador de grande importância, como Luisito Suarez, se posiciona no canto esquerdo da grande área, aguardando um cruzamento que pode resultar em gol, o treinador adversário deve desenvolver estratégias para assegurar que essa zona específica (canto esquerdo da grande área) seja coberta por um jogador da defesa.

Assim, a identificação dessas zonas de influência torna-se de extrema importância para compreender e antecipar táticas utilizadas pelos jogadores quando não estão em posse da bola. Isso permite uma análise mais abrangente e estratégica do comportamento dos atletas durante o jogo.

Assim, a proposta desse estudo é utilizar técnicas de estatística espacial na análise de dados de *tracking*, buscando identificar como a distância entre os jogadores que estão ou não estão em posse da bola podem sugerir ou caracterizar aspectos táticos de uma equipe. Especificamente, será utilizado um modelo geoestatístico (Cressie (1993)), que estima uma superfície aleatória em determinada região do espaço. No caso, a região do espaço será o campo de futebol e a superfície aleatória será a probabilidade de observar um jogador de determinado time em determinada posição. Cada uma dessas superfícies será estimada considerando diferentes ações dos jogadores. Espera-se que as estimativas das superfícies das probabilidades das posições dos jogadores para uma ação, como passe do atacante, seja bastante diferente do passe de um goleiro, indicando diferentes estruturas de dependência espacial.

1.4 Organização do trabalho

Este trabalho está organizado da seguinte forma: o capítulo 2 apresenta uma breve introdução à história do futebol, além de um breve apanhado histórico do uso da estatística no futebol e demais esportes (principalmente no contexto das ligas profissionais norte-americanas). No capítulo 3 apresentamos a metodologia de modelagem da pesquisa, mostrando o que é um processo espacial, o que é um modelo hierárquico e as técnicas bayesianas utilizadas na análise. No capítulo 4 aborda a forma como serão realizadas as análises, com a utilização do *software* estatístico **R** para a manipulação dos dados e para a construção do modelo espaço-temporal utilizaremos a biblioteca R-INLA, além de apresentarmos os resultados obtidos. No capítulo 5 apresentamos as considerações finais e, por fim, no capítulo 6 temos os anexos com as tabelas com todas os valores de das modas das distribuições *a posteriori* dos parâmetros estimados.

2 Futebol e Estatística

2.1 O futebol

Apesar de jogos que envolvam chutar um objeto esférico seja documentado ao longo da história em diversas culturas ao redor do mundo, a história da criação do futebol, como conhecemos, remonta ao século XIX, e é atribuída aos ingleses.

A prática de esportes era comum e incentivada nas escolas britânicas como forma de gerar camaradagem, competitividade e espírito de equipe. Entretanto, apesar de diversos jogos, cada escola tinha sua própria regra para a prática deles, o que dificultava o enfrentamento entre essas escolas em competições regionais.

Durante décadas se tentou chegar a um acordo sobre regras que seriam utilizadas e a discórdia na formulação de regras entre os participantes acarretou na cisão desses grupos e a criação de diversos esportes que conhecemos hoje, como o rugby. Até que em 1863, foi fundada a Football Association (Associação de Futebol), na Inglaterra, onde foi discutido e selado o primeiro conjunto de regras que deram início ao futebol como conhecemos. Essas regras ajudaram a padronizar o esporte e a criar uma base comum para sua prática.

O futebol rapidamente se espalhou pelo Reino Unido e, posteriormente, alcançou outros países através de expatriados, marinheiros e comerciantes britânicos. Em muitas nações, o esporte foi adotado e adaptado, ganhando popularidade ao redor do mundo.

No Brasil, o futebol foi introduzido por Charles Miller, um brasileiro de ascendência britânica, que estudou na Inglaterra. Em sua bagagem, trouxe bolas, uniformes e o conhecimento das regras do futebol. A popularidade do futebol cresceu rapidamente no Brasil, espalhando-se por clubes e escolas. O esporte não apenas oferecia uma atividade física, mas também proporcionava uma identidade cultural e social que o moldou como um dos esportes mais praticados no mundo.

Hoje, no país, de acordo com as informações da Confederação Brasileira de Futebol (CBF) divulgadas no ano de 2023, há 27 federações afiliadas e 1276 clubes registrados (Confederação Brasileira de Futebol - CBF (2023)). Devido ao grande número de praticantes e entusiastas do esporte e aos seus notáveis resultados alcançados em competições internacionais ao longo da história, o Brasil é reconhecido mundialmente como a 'nação do futebol' e como um lugar de muitas revelações de talentos excepcionais no cenário do futebol a nível mundial (Custódio (2011)).

2.2 A estatística

2.2.1 Estatística em outros esportes

Ao longo das últimas décadas, a estatística se estabeleceu como uma ferramenta essencial nos esportes, principalmente entre os americanos, transformando a maneira como equipes avaliam o desempenho de jogadores e tomam decisões estratégicas. O fenômeno, conhecido como "Moneyball", ganhou notoriedade especialmente na Major League Baseball (MLB), mas suas influências se estenderam para outras ligas profissionais.

Com a publicação do livro "Moneyball", de Michael Lewis, em 2003, que a estatística nos esportes ganhou destaque generalizado. O livro narra a história de Billy Beane, gerente geral do Oakland Athletics, que utilizou métodos estatísticos avançados para formar uma equipe competitiva com um orçamento limitado. Essa abordagem se concentrou em estatísticas como porcentagem de base e corridas criadas, proporcionando uma visão mais abrangente do valor de um jogador.

Desde então, equipes em várias ligas esportivas americanas adotaram métodos estatísticos avançados para ganhar vantagem competitiva. O basquete, por exemplo, viu a ascensão de métricas avançadas como o PER (*Player Efficiency Rating*), que avalia a eficiência global de um jogador em diversos aspectos do jogo.

De acordo com o renomado site de estatísticas de basquete, Basketball Reference, esta métrica foi criada por John Hollinger. Este é um sistema que atribui uma nota a todo instante a um jogador, em que o modelo soma todas as ações positivas, diminui as ações consideradas negativas e assim atribui um *score* para a perfomance do atleta (Basketball-Reference (2023)).

Algoritmos de análise de desempenho, aprendizado de máquina e inteligência artificial também têm sido empregados para prever resultados de jogos, identificar padrões de jogo e otimizar estratégias táticas. Isso cria uma abordagem mais científica para as decisões no esporte, indo além da intuição e da experiência.

Ainda na NBA a liga assinou parceria com a Microsoft para lançar uma plataforma chamada NBA CourtOptix (Wired (2020)). Esta plataforma utiliza *machine learning* e *AI* para analizar os movimentos dos jogadores na quadra, processar eles em tempo real utilizando o serviço em nuvem da empresa, o Azure, e entregar aos usuários métricas de análise mais robustas.

A cada temporada mais equipes que abraçam essa abordagem analítica e buscam maneiras inovadoras de avaliar talentos, melhorar o desempenho e alcançar o sucesso competitivo em um cenário esportivo cada vez mais desafiador.

2.2.2 Origens no futebol

Um dos primeiros registros que temos é de Charles Reep, conhecido como o pai do uso de estatística no futebol (Sykes e Paine (2016)). Nos anos 50, ele assumiu uma responsabilidade por conta própria: acompanhar todos os jogos do Swindon Town. Munido de um bloco de notas, ele anotava, na mão, todas as jogadas de todos os jogos que assistia, inclusive, desenhando o posicionamento de jogadores dentro de campo, 60 anos antes do surgimento das primeiras empresas que coletam esse tipo de informação.

Ao analisar os dados que ele mesmo coletara, Reep chegou a uma conclusão: a maioria dos gols numa partida de futebol são precedidos de 3 passes ou menos, ou

seja, o segredo para marcar mais gols era realizar passes longos para chegar à meta adversária o mais rápido possível. Apesar de se basear em uma lógica falaciosa, não podemos atribuir somente a ele o atraso dos ingleses no esporte nas próximas décadas.

2.2.3 Técnicas com bola

O que antes era apenas estatística descritiva passou a agregar técnicas mais complexas. O seu uso como ferramenta auxiliar no estudo e preparação do esporte se deu na virada para o século XXI. Com a maior acessibilidade e capacidade de análise de dados, pesquisadores passaram a ter uma gama maior de possibilidades de estudo.

A primeira grande mudança na forma de analisar dados no futebol foi com Sarah Rudd em 2011. Inspirada pelo *Santo Graal* da estatística no esporte, "Moneyball", ela apresentou em uma conferência em Harvard uma métrica nova. Hoje conhecida como *Expected Threat*, ela dividiu o jogo em diversos momentos e usou cadeias de Markov para calcular a probabilidade de mudança de um momento para outro (Sumpter (2017)).

Outra métrica com o mesmo nome foi desenvolvida por Karun Singh, hoje estatístico no Arsenal da Inglaterra. O modelo mais básico para avaliar a qualidade de uma posse de bola é baseada na localização. O hoje popular xG (*expected goals*) leva em consideração a posição do jogador que finalizou, distância para o gol, se havia marcadores no caminho, e outros parâmetros conforme o autor do modelo. Ao invés de focar apenas no momento da finalização (ou no passe imediatamente anterior), Singh buscou uma forma de calcular a probabilidade de se marcar um gol de acordo com o local que a bola se encontra.

Ao dividir o campo em zonas e calcular probabilidades de chute ou passe com base em histórico de dados em cada zona, a ideia é dar um peso a cada ação de forma a dar uma noção de "perigo" daquela posse de bola resultar em um gol, por isso o nome *Expected Threat* (Singh (2018)).

2.2.4 Técnicas sem bola

No artigo intitulado *Dynamic analysis of team strategy in professional football*, os autores Laurie Shaw e Mark Glickman, propuseram uma nova técnica para medir e classificar dinamicamente as formações de equipes em partidas de futebol profissional. Para tal, utilizaram uma grande amostra de dados de rastreamento de jogadores, em que foi medido o posicionamento relativo dos jogadores de cada equipe com e sem a posse da bola em sucessivos intervalos de tempo durante cada partida (Shaw e Glickman (2020)).

Outro artigo, publicado em conferência realizada pelo Massachusetts Institute of Technology (MIT), os autores Javier Fernandez e Luke Bornn apresentaram um método para quantificar a ocupação e a geração de *scores* de diferentes posições de um jogador durante uma partida. Os autores criaram um modelo para mensurar o controle da ocupação do campo, o qual incorpora informações de movimento, distância relativa à bola e posição do jogador. Tal modelo identifica áreas de ocupação de posse de bola, além disso o *score* de qualquer posição do campo, com relação a posição da bola é calculado através de redes neurais (Fernandez e Bornn (2018)).

Hoje em dia, o que antes eram esforços praticamente individuais na evolução de modelos explicativos para aprimorar a compreensão do futebol (e de outros esportes), agora faz parte do cenário corporativo. Empresas dedicadas a oferecer soluções para clubes e interessados na busca por uma interpretação mais aprofundada do jogo estão surgindo com frequência. Nomes como Statsbomb (Statsbomb (2021)) e Stats Perform (Stats-Perform (2019)), assim como blogs, como o American Soccer Analysis (Kullowatz (2020)), estão desenvolvendo modelos cada vez mais inovadores, aplicando diversos métodos com o objetivo de elevar o nível de análise disponível ao público.

3 Metodologia

3.1 Estatística espacial

Dada a natureza intrínseca de certos tipos de dados, é sensato empregar um modelo que leve em consideração a distância entre os pontos. Espera-se que a variação na saída desse modelo seja semelhante em regiões próximas, alinhando-se com a Primeira Lei da Geografia, que postula que "tudo está relacionado a tudo, mas coisas próximas têm uma relação maior do que as distantes" (Tobler (1970)).

O modelo a ser utilizado na análise das regiões de influência deve levar em consideração a distância entre os diferentes atletas das duas equipes. Uma área da estatística que envolve a análise desse tipo de dado é a estatística espacial, especificamente análise de dados de padrões pontuais (Gotway e Waller (2004)) e análise geoestatística , em que a região em estudo $\mathcal{D} \in \mathbb{R}^2$, representando alguma região mensurada de forma contínua (Cressie (1993)).

Em geoestatística e em padrões pontuais os dados consistem em informações medidas em locais específicos, sendo essas localizações definidas em um sistema de coordenadas, frequentemente expressas em longitude e latitude (Krainski et al. (2019)). No nosso contexto, o sistema de referência é estabelecido pela posição dos 22 jogadores em um campo de futebol durante uma partida.

Neste trabalho serão abordadas técnicas que envolvem geoestatística onde a localização dos casos observados não é aleatória, porém fixa em cada momento do jogo, além do tamanho da amostra ser previamente conhecido.

3.1.1 Modelo geoestatístico

Um exemplo de uso de um modelo geoestatístico é a estimação de precipitação mensal em uma região $\mathcal{D}$, a precipitação acontece de forma contínua, podendo ser observada em toda a área ($\mathcal{D} \in \mathbb{R}^2$), chamamos esse tipo de variável de um campo aleatório, como exemplificado na Figura 3.1 retirada do livro Krainski et al. (2019). Se desejamos prever a quantidade de chuva em uma área $\mathcal{D}$, é necessário distribuir estações meteorológicas que façam a mensuração da precipitação em determinadas localidades.

Utilizando as mensurações dessas estações, é possível estimar um modelo geoestatístico, que leve em consideração não apenas a distância entre as estações, mas também a variabilidade espacial da chuva na região. Com base nesse modelo, podemos fazer previsões mais precisas sobre a quantidade de chuva em locais onde não há uma estação meteorológica, utilizando as informações das estações próximas e a

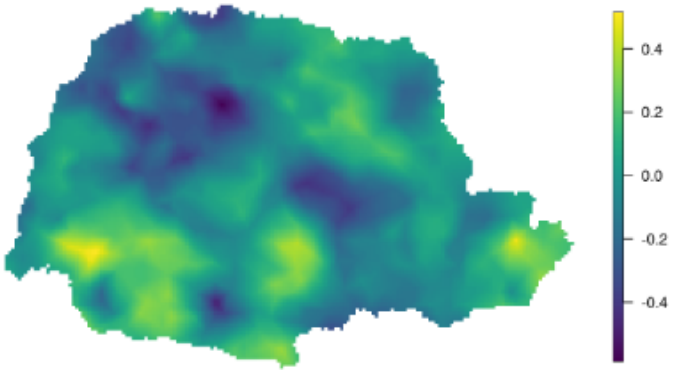


Figura 3.1: Gráfico de um campo aleatório, representando a média *a posteriori* da quantidade de precipitação no estado do Paraná.

estrutura de correlação espacial dos dados. Logo, o modelo estima uma superfície contínua de chuva na região de interesse, considerando uma quantidade discreta e limitada de estações de monitoramento.

Neste trabalho vamos aplicar um princípio parecido. As estações são os jogadores de ambas as equipes e a distância entre eles, no nosso caso as zonas de influência, vão depender se o jogador pertence ao time da casa ou ao time visitante.

Um processo espacial (ou campo aleatório) $\{y(s_{lat}, s_{long}) = y(\mathbf{s}), \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^{\mathbf{2}}\}$ é caracterizado por uma indexação espacial $\mathbf{s}$ que varia continuamente em uma área de estudo, que chamaremos de $\mathcal{D}$. Assim, $Y(\mathbf{s})$ é o processo estocástico, com $\mathbf{s} \in \mathcal{D}$ onde $\mathcal{D} \subset \mathbb{R}^2$.

O processo espacial é um campo gaussiano se para qualquer $n \geq 1$ e para cada um dos conjuntos de localizações $(\mathbf{s}_1, \ldots, \mathbf{s}_n)$, o vetor $\mathbf{y} = (y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, isto é, se segue uma distribuição Normal multivariada com média $\boldsymbol{\mu} = (y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n))$ e matriz de covariância espacialmente estruturada $\boldsymbol{\Sigma}$. A matriz $\boldsymbol{\Sigma}$ é definida pela função de covariância $\mathcal{C}(\cdot, \cdot)$ de modo que a matriz de covariância é dada por:

$$\Sigma_{ij} = Cov(y(\mathbf{s}_i), y(\mathbf{s}_j)) = \mathcal{C}(y(\mathbf{s}_i), y(\mathbf{s}_j)). \tag{3.1.1}$$

Assim, sendo $y(\mathbf{s}_i)$ uma realização de um processo espacial gaussiano, uma forma de representar um modelo geoestatístico é:

$$y(\mathbf{s}_i) = \mu + \epsilon(\mathbf{s}_i)$$

no qual $\mu = \beta_0 + \beta_1 X$ e o efeito aleatório $\epsilon(\mathbf{s}_i)$ possui algum tipo de estrutura espacial, indicando que observações próximas são mais similares que observações distantes. Caso o vetor $\boldsymbol{\epsilon}$ tenha distribuição Normal multivariada, como acima, então tem-se que o vetor $\boldsymbol{\epsilon}$ é um campo gaussiano com vetor de médias $\boldsymbol{\mu} = 0$e matriz de covariância dada pela Equação (3.1.1).

Algumas propriedades importantes de um campo aleatório estão associadas à estacionariedade de primeira e de segunda ordem. Se um processo espacial é estacionário de primeira ordem, então a função média é constante no espaço, ou seja, $E[y(\mathbf{s}_i)] = \mu$ para qualquer i. Se um campo aleatório é estacionário de segunda ordem, então além de $E[y(\mathbf{s}_i)] = \mu$, $\forall \quad \mathbf{s}_i \subset \mathcal{D}$, a função de covariância espacial depende apenas da distância $(\mathbf{s}_i - \mathbf{s}_j)$, ou seja, $Cov(y(\mathbf{s}_i), y(\mathbf{s}_j)) = \mathcal{C}(||y(\mathbf{s}_i) - y(\mathbf{s}_j)||)$.

Ademais, caso a função de covariância não dependa da direção, mas apenas da distância euclidiana entre os pontos, o processo é chamado de isotrópico (Blangiardo e Camelatti (2015)).

Dessa forma, ao assumir estacionariedade de segunda ordem e isotropia em um processo espacial, ganha-se vantagens significativas em termos de simplicidade e interpretação, permitindo uma análise mais simplista que pode ser adequada mesmo para modelos complexos.

3.1.2 Covariância de Matérn

Ao considerar a modelagem de um processo espacial, a escolha adequada da função de covariância desempenha um papel crucial na captura eficaz das relações entre as observações ao longo do espaço. Em geoestatística a função de covariância de Matérn é comumente utilizada, devido às suas propriedades flexíveis e capacidade de capturar diferentes níveis de suavidade nas variações espaciais dos dados.

A flexibilidade da função de covariância de Matérn a torna adequada para lidar com diferentes padrões de variabilidade, desde variações mais suaves até variações mais abruptas. Isso é especialmente útil em modelos espaciais, nos quais as características dos dados podem variar significativamente ao longo do espaço (Moraga (2019)).

Assim, ela é uma escolha popular nesse tipo de modelagem, devido à sua capacidade de se adaptar a diferentes padrões de variabilidade, proporcionando uma representação robusta e flexível do processo espacial em análise.

A função de covariância de Matérn é definida por:

$$Cov(Y_i, Y_j) = Cov(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) = \frac{\sigma^2}{\Gamma(\lambda)2^{\lambda-1}}(\kappa||\mathbf{s}_i - \mathbf{s}_j||)^{\lambda} K_{\lambda}(\kappa||\mathbf{s}_i - \mathbf{s}_j||)$$

em que $||\mathbf{s}_i - \mathbf{s}_j||$ é a distância Euclidiana entre duas localidades $\mathbf{s}_i, \mathbf{s}_j \in \mathbb{R}^2$, σ^2 é a variância marginal do campo Gaussiano, K_{λ} denota a função de Bessel modificada de ordem λ, λ mede o grau de suavidade do processo e $\kappa > 0$ é um parâmetro conhecido como *range*, que indica a partir de qual distância a correlação espacial se torna quase nula (Blangiardo e Camelatti (2015)).

3.2 Modelos Hierárquicos Bayesianos

Como já mencionado, neste trabalho pretendemos abordar como os jogadores influenciam com e sem a posse da bola. Porém, um jogo de futebol não é estanque: os jogadores constantemente alteram suas posições, movem-se em diversas direções e, por vezes, modificam completamente suas trajetórias de acordo com as ações dos companheiros de equipe e o movimento da bola. Assim, precisamos adicionar um novo elemento dentro da análise, a variação temporal. Nesse sentido, uma abordagem é a aplicação de modelos hierárquicos bayesianos, os quais se revelam eficazes na modelagem de dados espaço-temporais (Moraga (2019)).

Em um contexto bayesiano, seja $p(\mathbf{y}|\boldsymbol{\theta})$ a distribuição de probabilidade dos dados $\mathbf{y} = (y_i, \ldots, y_n)$, dado um vetor de parâmetros desconhecidos $\boldsymbol{\theta}$, sabemos que:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}, \boldsymbol{\theta})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\mathbf{y})}{\int p(\mathbf{y}|\boldsymbol{\theta})p(\mathbf{y})d\boldsymbol{\theta}},$$

na qual o denominador define a verossimilhança marginal dos dados $\mathbf{y}$. Como esta verossimilhança não depende do parâmetro $\boldsymbol{\theta}$, ela pode ser entendida como uma constante que não influencia no resultado da distribuição *a posteriori*, de forma que esta distribuição pode ser escrita como:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\mathbf{y}).$$

Uma vantagem da metodologia bayesiana é a possibilidade de incorporar um conhecimento prévio ao modelo e assim atualizar a informação prévia baseado na observação dos dados. Além disso, métodos bayesianos possibilitam a incorporação de modelos mais complexos que seriam difíceis de modelar usando a estatística clássica. Como queremos incorporar o elemento tempo dentro da modelagem, esta é uma técnica adequada ao que pretendemos modelar (Blangiardo e Camelatti (2015)).

Importante ressaltar que, em um contexto bayesiano, a interpretação dos parâmetros difere da estatística clássica, uma vez que estes são caracterizados por uma distribuição de probabilidade. Ao contrário da teoria clássica, na qual os parâmetros são considerados desconhecidos e fixos, no enfoque bayesiano, sua representação é mais flexível, refletindo a incerteza associada a esses valores por meio de uma distribuição probabilística (Blangiardo e Camelatti (2015)).

3.3 Modelagem

3.3.1 Modelo proposto

A modelagem visa acompanhar a movimentação dos 22 jogadores e verificar o quanto da distância entre eles é relativa a seus companheiros de time e o quanto é devido ao tempo de jogo. O modelo a ser considerado será o binomial, em que sucesso irá significar o time que é da casa e fracasso se refere ao time visitante, dado o tipo de jogada. As posições dos jogadores serão conhecidas em todos os *frames* e o que será aleatório é o time ao qual o jogador pertence. Nesse caso, a superfície aleatória a ser estimada é a probabilidade do jogador ser do time da casa.

Assim, seja $Y_{i,t}$ a variável binomial da i-ésima localização espacial no tempo t. Podemos expressar o número de sucessos (ou eventos de interesse) em $n_{i,t}$ tentativas como uma variável aleatória binomial com probabilidade $p_{i,t}$, tal que $0 \leq p_{i,t} \leq 1$.

Podemos então definir o modelo espacial-temporal da seguinte forma:

$$Y_{i,t} \sim \text{Binomial}(n_{i,t}, p_{i,t}),$$

em que $n_{i,t}$ é o número total de eventos em $Y_{i,t}$ e $p_{i,t}$ é a probabilidade de sucesso para a i-ésima localização espacial no tempo t.

A relação entre a probabilidade de sucesso $p_{i,t}$ e os preditores pode ser modelada usando uma função de ligação. Vamos chamar $\eta_{i,t}$ de preditor linear, logo, a relação entre $p_{i,t}$ e $\eta_{i,t}$ é dada por uma função de ligação $g(\cdot)$. Uma escolha comum é a função logit para dados binomiais (Blangiardo e Camelatti (2015)):

$$g(p_{i,t}) = \ln\left(\frac{p_{i,t}}{1-p_{i,t}}\right) = \eta_{i,t}.$$

Agora, para incorporar a estrutura espacial e temporal, podemos incluir efeitos aleatórios espaciais e temporais. Suponhamos que $\xi(\mathbf{s}, t)$ represente o efeito aleatório espacial-temporal associado a uma localização espacial $\mathbf{s}$ no tempo t.

A equação do modelo espacial-temporal binomial seria então:

$$\eta_{i,t} = g(p_{i,t}) = \beta_0 + \beta_1 X_{i,t} + \xi(\mathbf{s}_i, t), \tag{3.3.1}$$

em que β_0 é o intercepto, β_1 é o coeficiente associado ao preditor $X_{i,t}$, e $\xi(\mathbf{s}_i, t)$ é o termo de efeito aleatório espaço temporal.

Nesta equação a variável resposta é modelada como uma distribuição binomial e a estrutura espaço temporal é incorporada através do termo de efeito aleatório $\epsilon(\mathbf{s}_i, t)$. A inclusão dos efeitos aleatórios é crucial para lidar com a complexidade dos dados uma vez que os pontos (jogadores) estão mudando de posição ao longo da partida. Os efeitos espaçotemporais ajudam a considerar a autocorrelação espacial (dependência espacial) e temporal (dependência temporal) nos dados, permitindo uma modelagem mais realista e precisa.

Existem várias abordagens para modelar a estrutura espaço temporal dos dados, e essas estruturas podem ser classificadas em funções de covariância separáveis e não separáveis Sherman (2011). No contexto deste trabalho, optamos por adotar um modelo com estrutura separável. Isso implica que a Equação (3.3.1) será decomposta em duas partes distintas, uma que considera a estrutura espacial e outra que captura a dependência temporal:

$$g(p_{i,t}) = \beta_0 + \beta_1 X_{i,t} + \epsilon(\mathbf{s}_i) + \delta_t, \tag{3.3.2}$$

em que β_0 é o intercepto, β_1 é o coeficiente associado ao preditor $X_{i,t}$, $\epsilon(\mathbf{s}_i)$ é o efeito aleatório espacial e δ_t é o efeito aleatório temporal, que será modelado conforme um autoregressivo de ordem 1 e parâmetro ρ.

Para a parte espacial do modelo já definimos que iremos utilizar a estrutura de covariância de Matérn. Para a parte temporal do modelo, utilizaremos a estrutura "AR1", que assume que a correlação entre as observações decresce exponencialmente à medida que a distância temporal entre elas aumenta. Isso significa que observações mais próximas no tempo terão uma correlação mais alta do que aquelas mais distantes.

3.3.2 Técnicas computacionais utilizadas

O modelo parte de duas premissas: por conta da variável resposta (o número de jogadores que fazem parte do time do portador da bola), ser binária, é adequado o uso de um modelo binomial. Além disso, se desejarmos incluir na modelagem variáveis preditoras, é comum tratar essa estrutura de dependência como um modelo linear, utilizado quando o valor esperado da variável resposta é uma função linear das covariáveis. Nessa caso, o modelo faz parte de uma classe denominada modelos lineares generalizados Nelder e Wedderburn (1972). Ademais, por se tratar de dados espaciais, além do preditor linear é necessário incluir a estrutura de dependência espacial no modelo.

Uma vez que lidamos com um modelo geoestatístico, é possível operar sob a suposição de que o efeito aleatório segue uma distribuição normal multivariada, e, portanto, torna-se viável empregar um campo Gaussiano para representá-lo de maneira adequada. Como a função da média será estimada através de um modelo linear

generalizado, os modelos hierárquicos bayesianos surgem como uma ótima alternativa. Na modelagem bayesiana é necessário computar as distribuições *a posteriori* dos parâmetros, quando as distribuições *a priori* são conjugadas com a verossimilhança as distribuições *a posteriori* possuem forma fechada Lee (2012). Quando isso não ocorre, que é o caso do modelo empregado nesse trabalho, é necessário empregar métodos numéricos, como o conhecido método de Monte Carlo via Cadeias de Markov (MCMC) Lunn et al. (2000).

Entretanto, para realizar a estimação de modelos que utilizam o método MCMC (Moraga (2019)) é necessário avaliar a convergência da cadeia, o que pode demandar um grande esforço computacional se a quantidade de estimativas for grande. Logo, um método que não exige esse tipo de cuidado pode ser mais rápido e mais confiável.

Assim, uma opção ao método de Monte Carlo via Cadeias de Markov (MCMC) é a aplicação do método INLA (Rue et al. (2011)).

3.3.3 SPDE

As Equações Diferenciais Parciais Estocásticas (SPDE, do inglês *Stochastic Partial Differential Equations*) são extensões das equações diferenciais parciais clássicas que incorporam componentes estocásticos, ou seja, incluem aleatoriedade no modelo. Enquanto as equações diferenciais parciais tradicionais descrevem a evolução de fenômenos determinísticos, as SPDEs lidam com sistemas influenciados por processos estocásticos, como o ruído aleatório.

Em 2011, Lindgren, Rue, and Lindström (Rue et al. (2011)) descreveram uma aproximação de modelos espaciais contínuos com covariância de Matérn baseada em uma solução para uma equação diferencial estocástica parcial. Nessa abordagem, um processo espacial contínuo $Y(\mathbf{s})$ é representado utilizando a função de covariância de Matérn e um processo aleatório espacial indexado discretamente, conhecido como Campo Aleatório Markoviano Gaussiano (GMRF). Ao contrário de um campo gaussiano, que possui uma matriz de covariância "cheia", um campo markoviano é caracterizado por uma matriz de covariância esparsa, proporcionando vantagens computacionais significativas, tornando o cálculo de inversas e determinantes muito mais rápido.

Como o SPDE transforma um campo gaussiano em um campo markoviano, a estrutura de dependência espacial que é representada através de uma distribuição normal multivariada, logo a incerteza é um modelo gaussiano latente, o que possibilita que o cálculo da distribuição *a posteriori* possa ser efeituado através do INLA.

3.3.4 R-INLA

O INLA (*Integrated Nested Laplace Approximation*), foi desenvolvido por Havard Rue, Sara Martino e Nicolas Chopin e apresentado no *Journal of the Royal Statistical Society - B* (Rue et al. (2009)). Trata-se de um algoritmo determinístico que combina técnicas de integração numérica e aproximação Laplaciana para calcular a distribuição *a posteriori* de modelos bayesianos. O INLA é apropriado especialmente na estimação de modelos gaussianos latentes. Particularmente em contextos espaço-temporais, é comum representar os efeitos aleatórios espaciais e temporais através de uma distribuição normal, logo o INLA é um método apropriado para obtenção das distribuições *a posteriori*. Além disso, em comparação com o MCMC,

o INLA oferece resultados precisos em um intervalo de tempo mais curto.

Para os cálculos das estimativas, será utilizado o pacote do *software* **R** R Core Team (2024) denominado R-INLA R-INLA Project (2024). Esse pacote é usado para realizar análises bayesianas de modelos hierárquicos complexos. O R-INLA é especialmente útil para lidar com modelos espaciais, temporais e mistos.

4 Resultados

4.1 Banco de dados

Nos esportes, "dados de rastreamento" referem-se a dados espaço temporais de granulação fina que descrevem as posições da bola e/ou do jogador durante um evento esportivo. Os dados de rastreamento de esportes modernos normalmente incluem coordenadas espaciais 2D de localizações de jogadores e coordenadas espaciais 3D de localizações de bolas em uma frequência de amostragem de 25 *Hertz* ou mais Kovalchik (2022).

O tipo de dado utilizado nesse estudo ainda é bastante complexo de conseguir. Ao contrário de *event data*, que se tem acesso relativamente fácil e gratuito a grandes bases de dados e dependendo da competição com atualizações em tempo real, *tracking data* não tem acesso fácil. Após extensa pesquisa e tentativas de entrar em contato com algumas equipes e profissionais que atuam na área, apuramos que existem apenas alguns repositórios online com esse tipo de dado.

Metrica Sports é uma empresa holandesa que vende diversos serviços relacionados a dados em esportes, sendo um destes, a coleta, o armazenamento e tratamento de *tracking data* (Metrica-Sports (2024)). Os dados deste repositório foram ofertados em parceria com David Sumpter, autor do livro *Soccermatics* e organizador do curso *Friends of Tracking* e informações relativas a times, jogadores, campeonato, entre outros dos 3 jogos disponibilizados são anônimos de forma que não podemos identificar qualquer jogador que está em campo.

Skillcorner é uma empresa sediada na França que tem como produto principal trabalhar com *video tracking* e análise de dados de esportes (Skillcorner (2024)). Os 9 jogos disponibilizados são *tracking data* obtidos através de imagens de transmissão de jogos. Esses jogos são da temporada 2019/2020 entre os campeões e vice-campeões das 5 principais ligas da Europa: Premier League inglesa, a Ligue 1 francesa, a LaLiga espanhola, a Serie A italiana e a Bundesliga alemã.

Assim, estamos limitados a apenas algumas partidas que estão disponíveis nos seguintes repositórios:

- `https://github.com/metrica-sports/sample-data`: da Metrica-Sports com 3 jogos;

- `https://github.com/SkillCorner/opendata`: da Skillcorner com 9 jogos.

Devido a limitação que a transmissão de jogos gera (como não ter todos os atletas na tela ao mesmo tempo), optamos por utilizar os jogos da Metrica-Sports.

Por se tratar de um modelo complexo que demanda muito desempenho computacional, para este trabalho resolvemos rodar o modelo em apenas um jogo. O

jogo escolhido foi o *Sample Game 1* que acabou 3x0 para o time da casa e teve as seguintes estatísticas:

Estatística	Time Casa	Time Fora
Gols	3	0
Chutes	18	6
Chutes no alvo	7	3
Chutes fora	10	3
Chutes bloqueados	1	0
Escanteios	8	3
Impedimentos	0	1
Faltas sofridas	15	7
Cartões amarelos	2	2
Passes certos	437	362

Tabela 4.1: Estatísticas do jogo *Sample Game 1*

O repositório do jogo tem 3 bancos de dados distintos em formato *.csv*: *tracking* do time da casa, *tracking* do time visitante, e um terceiro banco com o *event data* sincronizado com os outros dois.

Os dois bancos de *tracking data* possuem a seguinte estrutura:

- *Period* - Tempo do jogo (se é primeiro ou segundo tempo)
- *Start Frame* - Frame que foi capturado
- *Start Time [s]* - Tempo em segundos que foi capturado
- *Player1* - Coordenada no "eixo x" do jogador 1
- *Player1 y* - Coordenada no "eixo y" do jogador 1
- $\vdots$
- *Player14* - Coordenada no "eixo x" do jogador 14
- *Player14 y* - Coordenada no "eixo y" do jogador 14

Os dados são coletados numa frequência de 25 *frames* por segundo, de forma que ao todo temos 145006 linhas para esse jogo. As dimensões do campo são 105x68 metros (ver Figura 4.1), porém os dados foram coletados de forma que em cada eixo varia de 0 a 1, com a coordenada (0,0) sendo o canto inferior esquerdo, (1,1) sendo o canto superior direito e (0.5,0.5) sendo o centro do campo. Para efetuar as análises, multiplicamos as coordenadas x e y por 105 e 68, respectivamente.

Em ambos os bancos existem registro de 14 jogadores que correspondem aos 11 jogadores que iniciaram a partida e também os 3 jogadores que eventualmente entraram como substitutos. Os dados referentes ao árbitro foram retirados.

Na Figura 4.2, temos o posicionamento dos jogadores nos primeiros 10 segundos de jogo. Em azul estão os jogadores do time da casa e em vermelho estão os jogadores do time de fora. A saída de bola foi dada pelo time visitante (em vermelho).

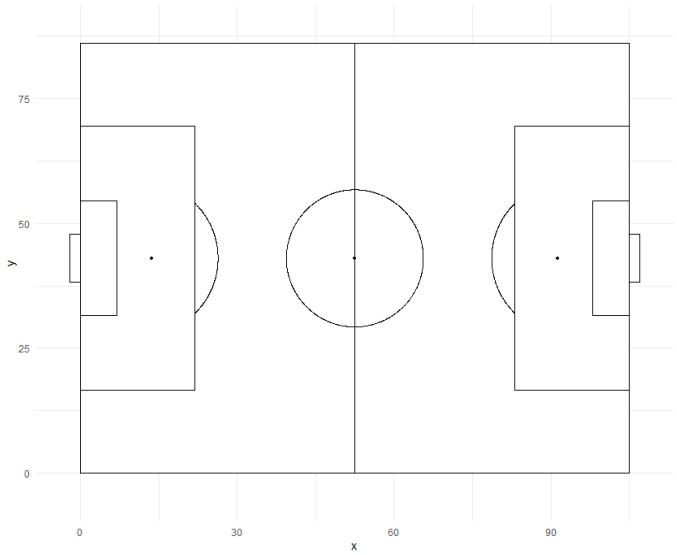
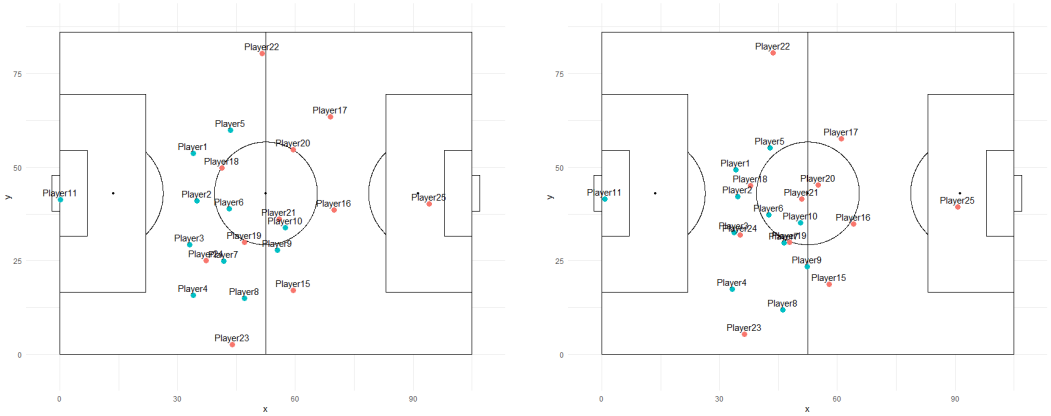
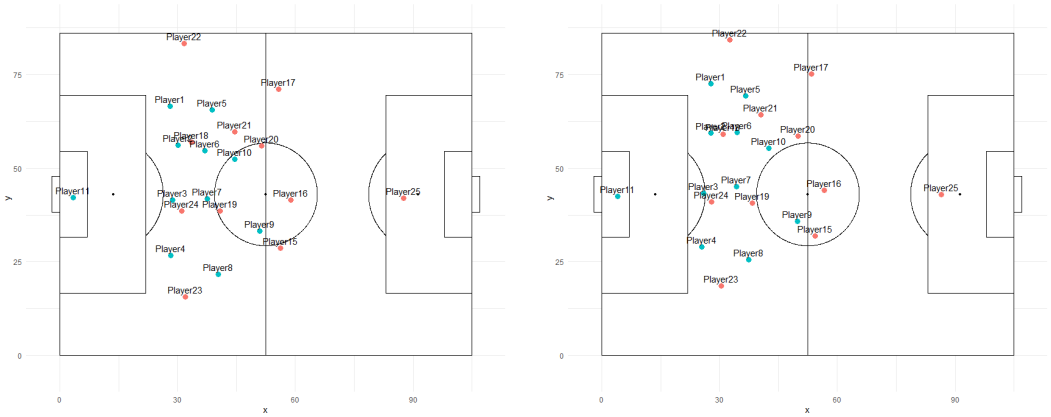


Figura 4.1: Campo de jogo com as dimensões projetadas



(a) *1s*

(b) *4s*



(c) *7s*

(d) *9s*

Figura 4.2: Posição dos jogadores nos tempos 1, 4, 7 e 9 segundos de jogo

No banco de *event data* cada linha representa uma observação de algum evento que ocorreu na partida, no total temos 1745 eventos, além disso, o banco de dados possui as seguintes 14 colunas:

- *Team* - Time
- *Type* - Tipo de evento
- *Subtype* - Subtipo de evento
- *Period* - Tempo do jogo (se é primeiro ou segundo tempo)
- *Start Frame* - Frame que o evento foi começa
- *Start Time [s]* - Tempo em segundos que o evento começa
- *End Frame* - Frame que o evento encerra
- *End Time [s]* - Tempo em segundos que o evento encerra
- *From* - Jogador que deu início ao evento
- *To* - Jogador que encerrou o evento
- *Start X* - Coordenada no "eixo x" que começou o evento
- *Start Y* - Coordenada no "eixo y" que começou o evento
- *End X* - Coordenada no "eixo x" que encerrou o evento
- *End Y* - Coordenada no "eixo y" que encerrou o evento

A variável *Type* está dividida em 9 tipos de eventos: *Set Piece, Pass, Ball Lost, Recovery, Challenge, Ball Out, Shot, Fault Received, Card.*

A variável *Subtype* está dividida em 24 subtipos de eventos: *Kick Off, NA, Interception, Head-Interception, Aerial-Lost, Aerial-Won, Head, Corner Kick, Cross, Head-On Target-Goal, Throw In, Ground-Lost, Ground-Won, Tackle-Won, Theft, Tackle-Lost, Off Target-Out, Goal Kick, Cross-Interception, Goal Kick-Interception, Deep Ball, On Target-Saved, Saved, Head-Forced, Aerial-Fault-Lost, Aerial-Fault-Won, Free Kick, Head-Clearance, Clearance, Ground-Fault-Won, Ground-Fault-Lost, Head-Off Target-Out, Dribble-Won, Offside, Tackle-Fault-Lost, Tackle-Fault-Won, Ground, Forced, Yellow, Through Ball-Deep Ball, Head-On Target-Saved, Blocked, Tackle-Advantage-Lost, Tackle-Advantage-Won, Ground-Advantage-Lost, End Half, Head-Woodwork-Out, Woodwork-Goal, On Target-Goal, Woodwork, Referee Hit, Off Target.*

Importante frisar que não é efetuado o rastreamento da bola. Dessa forma, para as análises foi considerado como portador da bola o jogadore que foi indicado que efetuou a ação registrada naquele instante.

Para maiores informações sobre como são classificados cada tipo de evento e subtipo de evento, basta acessar o link do repositório e acessar a página sobre a documentação do banco.

Team	Type	Subtype	Period	Start Frame	Start Time [s]	End Frame	End Time [s]	From	To	Start X	Start Y	End X	End Y
Away	SET PIECE	KICK OFF	1	1	0.04	0	0.00	Player19	NA	NaN	NaN	NaN	NaN
Away	PASS	NA	1	1	0.04	3	0.12	Player19	Player21	0.45	0.39	0.55	0.43
Away	PASS	NA	1	3	0.12	17	0.68	Player21	Player15	0.55	0.43	0.58	0.21
Away	PASS	NA	1	45	1.80	61	2.44	Player15	Player19	0.55	0.19	0.45	0.31
Away	PASS	NA	1	77	3.08	96	3.84	Player19	Player21	0.45	0.32	0.49	0.47
Away	PASS	NA	1	191	7.64	217	8.68	Player21	Player22	0.40	0.73	0.32	0.98
Away	PASS	NA	1	279	11.16	303	12.12	Player22	Player17	0.39	0.96	0.49	0.98
Away	BALL LOST	INTERCEPTION	1	346	13.84	380	15.20	Player17	NA	0.51	0.97	0.27	0.75
Home	RECOVERY	INTERCEPTION	1	378	15.12	378	15.12	Player2	NA	0.27	0.78	NaN	NaN
Home	BALL LOST	INTERCEPTION	1	378	15.12	452	18.08	Player2	NA	0.27	0.78	0.59	0.64
Away	RECOVERY	INTERCEPTION	1	453	18.12	453	18.12	Player16	NA	0.57	0.67	NaN	NaN
Away	BALL LOST	HEAD-INTERCEPTION	1	453	18.12	497	19.88	Player16	NA	0.57	0.67	0.33	0.65
Away	CHALLENGE	AERIAL-LOST	1	497	19.88	497	19.88	Player18	NA	0.38	0.67	NaN	NaN
Home	CHALLENGE	AERIAL-WON	1	498	19.92	498	19.92	Player2	NA	0.36	0.67	NaN	NaN
Home	RECOVERY	INTERCEPTION	1	498	19.92	498	19.92	Player2	NA	0.36	0.67	NaN	NaN

Figura 4.3: Imagem da tela do Rstudio com as primeiras linhas e todas as colunas do banco de dados dos eventos.

Para ter uma compreensão de como esses dados são formatados, na Figura 4.3 temos uma demonstração de como foram salvas as informações referente aos eventos da partida.

Como os dados de rastreamento são capturados em 25 *frames* por segundo, por uma questão de economia de processamento dos dados e entendendo que não há necessidade de um refino tão grande para capturar a presença de correlação temporal, optamos por fazer as análises segundo a segundo.

Assim, através das informações que constam no banco de eventos, calculamos as janelas de tempo de cada evento. Esse comprimento de "janela" foi calculada apenas com a subtração do *frame* final com o *frame* inicial de cada evento. Na Figura 4.4 temos o histograma de quantas ocorrências de cada tamanho de janela temos na partida. Retiramos do gráfico as janelas de valor 0 pois eram a grande maioria dos casos. Estas janelas correspondem a ação que são inerentemente instantâneas, como uma interceptação, por exemplo.

Através dessa imagem é possível observar que a maioria dos eventos possui uma baixa quantidade de *frames*. Desconsiderando as janelas de tamanho 0, a moda seria janelas de 30 *frames* com uma frequência total de 62 ocorrências. Analisando as informações geradas pela criação dessa janela de ações, temos que a mediana é referente a janela de ação de 20 *frames* e o primeiro e terceiro quartis são 0 e 37 *frames*, respectivamente.

Rodar o modelo para uma partida completa demanda muito tempo e custo computacional. Por isso, apresentamos os resultados das análises das ações de alguns jogadores, todos do time mandante, para compreender como eles influenciam o posicionamento dos demais. Foram escolhidos os atletas: *Player1* que imaginamos ser um lateral, *Player2* que imaginamos ser um zagueiro, *Player9* que imaginamos ser um atacante, *Player10* que imaginamos ser um meio-de-campo e *Player11* que é o goleiro.

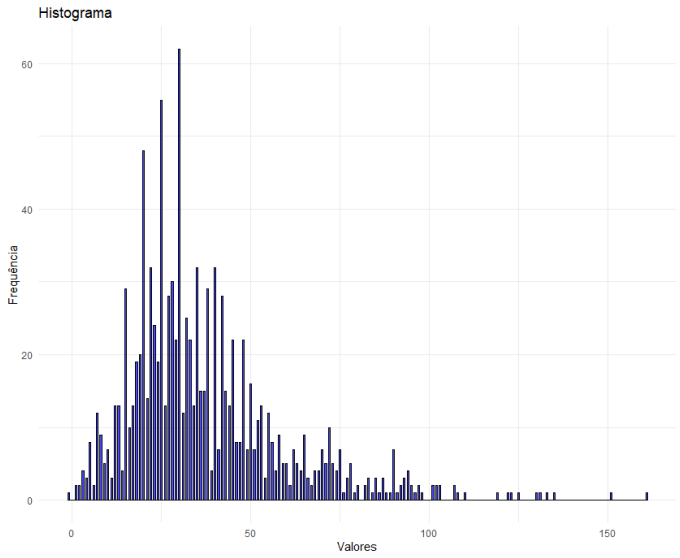


Figura 4.4: Histograma das janelas de *frames* de cada evento no jogo

4.2 Modelo gerado

O primeiro passo foi realizar a preparação dos dados de forma que estes sejam compatíveis com a modelagem que será efetuada. Como o nosso interesse é verificar se os jogadores próximos fazem parte do mesmo time, precisamos que o banco a ser utilizado tenha essas características presentes. Após algumas manipulações, o banco final possui as seguintes variáveis:

- *jogador* - Qual o jogador
- *coord x* - Coordenada no "eixo x" do jogador em questão
- *coord y* - Coordenada no "eixo y" do jogador em questão
- *home team* - Variável dicotômica, sendo 1 caso positivo e 0 caso contrário
- *away team* - Variável dicotômica, sendo 1 caso positivo e 0 caso contrário
- *frame* - Frame que o posicionamento dos jogadores foi capturado

4.2.1 Escolha do mesh do modelo SPDE

A abordagem SPDE, método que será utilizado na análise dos dados, aproxima o campo gaussiano contínuo como um campo markoviano. Assim, a área que antes era contínua é discretizada e o *mesh* representa este novo espaço, conforme se observa na Figura 4.5. A seguir, na Figura 4.6, temos o mesmo *mesh* porém com o campo de jogo sobreposto e as posições dos jogadores no intervalo de tempo referido. Em azul temos o time que joga em casa e em vermelho o time visitante.

A ideia da criação de um *mesh* é um *trade off* entre a precisão da representação de um campo gaussiano e o custo computacional, em que ambos dependem do número de vértices usados na triangulação: quanto maior o número de triângulos, mais precisa vai ser a aproximação do campo gaussiano porém os custos computacionais serão maiores.

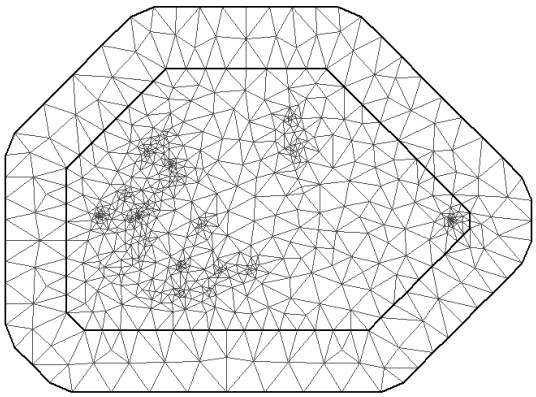


Figura 4.5: *Mesh* relativo aos segundos 3954 a 3961 da partida

Dentro do R-INLA existe uma função que monta o *mesh* e a partir dela podemos gerar uma matriz de projeção que será utilizada para atribuir os efeitos espaciais e temporais no modelo.

Para a construção do *mesh* definimos que o comprimento máximo dos lados dos triângulos a serem gerados na região seria de 8 e na área fora da região onde estão presentes os dados de 10 metros e assim não há desperdício computacional para o cálculo dessa área. Por exemplo, na Figura 4.5 temos um total de 790 vértices e a quantidade de vértices a serem montados depende da proximidade dos pontos.

4.2.2 Definição do modelo no pacote INLA

Definido o *mesh* passamos a construir o modelo SPDE, para isso é necesário especificar a distribuições *a priori* dos parâmetros do campo dado a função de covariância Matérn, na modelagem o parâmetro que controla a suvidade será fixo $\lambda = 2$

(κ, σ) para a dependência espacial e um auto regressivo de ordem 1 (ρ) para a dependência temporal, para isso iremos utilizar o método PC prior, *Penalised Complexity* Simpson et al. (2017).

Dentro do objeto precisamos definir, além do *mesh*, o *alpha*, que é relacionado a suavização do processo e que por *default* é igual a 2. Além disso, precisamos definir os parâmetros das PC *priores*. Para *range* do processo definimos que a probabilidade do processo espacial ter uma *range* maior do que 5 metros é próxima de 0.1. Para *sigma* (*marginal standard deviation*), que é referente a variabilidade dos dados, definimos que a probabilidade da variância ser maior que 6 é 0.1.

Após construímos um objeto que cria a indexação do modelo e a matriz de projeção. A indexação serve para definir o tipo de modelagem e quais efeitos serão incorporados ao modelo. A matriz de projeção projeta o campo gaussiano através das observações dos vértices do *mesh*. No presente trabalho o nosso foco principal foi estimar a distribuição *a posteriori* dos parâmetros acima mencionados, porém,

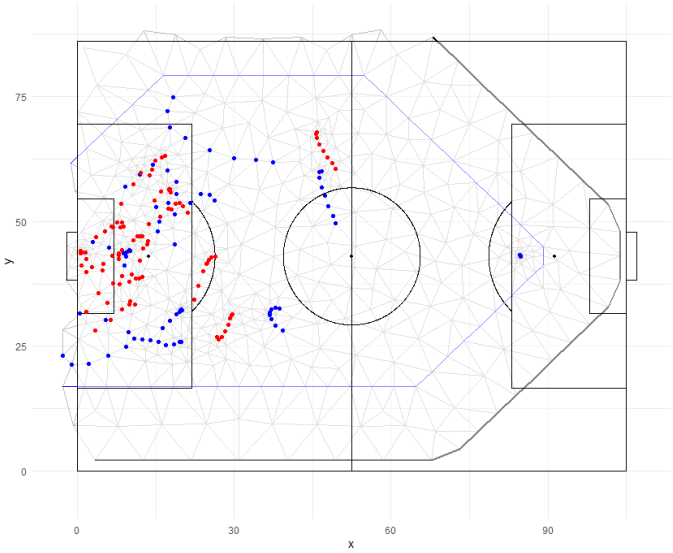


Figura 4.6: Campo de jogo sobreposto ao *mesh* relativo aos segundos 3954 a 3961 da partida

dependendo do tipo de modelagem, pode-se criar matrizes de projeção para além da estimação calcular a predição.

O próximo passo é construir os *stacks* para estimação (e predição, caso seja do interesse). Aqui atribuímos os tipos de efeitos que serão estimados para cada tipo de variável. Como mencionado, no presente trabalho estamos apenas interessados na estimação dos parâmetros espaço-temporais.

Definimos a distribuição *a priori* do parâmetro temporal do modelo como "pc-cor1", em que novamente utilizamos uma PC prior, em que assumimos que a probabilidade do parâmetro ρ ser maior que 0 é igual a 0.9. Por fim, precisamos definir a fórmula do modelo e é nesse momento que incluímos a correlação temporal. Como foi delimitado na definição do parâmetro, para este estudo definimos que a correlação espacial é da ordem de um "AR1".

4.2.3 Resultados

Como iremos demonstrar, os resultados gerados através dessa modelagem nos permite elaborar diversas conclusões acerca das relações entre os jogadores.

Por uma questão de brevidade do trabalho, o nosso foco neste momento foi verificar a influência espaço-temporal das ações de alguns atletas em relação aos demais. Por se tratar de dados sem identificação dos jogadores, observamos o posicionamento de alguns atletas, como na Figura 4.2, para inferir sobre a posição de cada um deles.

Através dos eventos dos jogos, filtramos as ações cuja origem eram dos jogadores *Player1, Player2, Player9, Player10 e Player 11* e rodamos o modelo na janela de ação correspondente a cada uma delas, para entender como as posições (aparentes) de cada um deles influencia no jogo.

No Anexo A são apresentadas as tabelas completas com as modas das distribuições *a posteriori* calculadas para os parâmetros do modelo, o parâmetro temporal ρ e os parâmetros da função de covariância matern κ e σ^2 para cada uma das ações dos atletas *Player1, Player2, Player9, Player10 e Player 11* .

4.3 Interpretações

Como é esperado, dependendo da posição do jogador e do tipo de ação, os valores da influência espacial e temporal mudam drasticamente. Por exemplo, o *Player9*, que, em tese, é um atacante, tem valores da *range* baixos em situações de chute (ver Tabela 5.3), o que indica uma proximidade maior de jogadores perto dele, e é razoável de se imaginar essa situação em um contexto de finalização. A sequência de *frames* do momento do gol (evento 19 do atleta) corrobora com essa ideia (Figura 4.7).

Na tabela 4.2, temos as médias das modas *a posteriori* computadas para cada um dos parâmetros considerando alguns dos tipos e subtipos de eventos referentes a ações realizadas pelos atletas. Como se observa, ações como finalizações que geraram gol (*Shot on target-goal*) possuem menor dependência espacial que um tiro de meta (*Pass goal* kick), o que é coerente com o imaginado uma vez que em uma finalização é esperado que mais jogadores estejam agrupados ao redor do jogador que vai efetuar o chute.

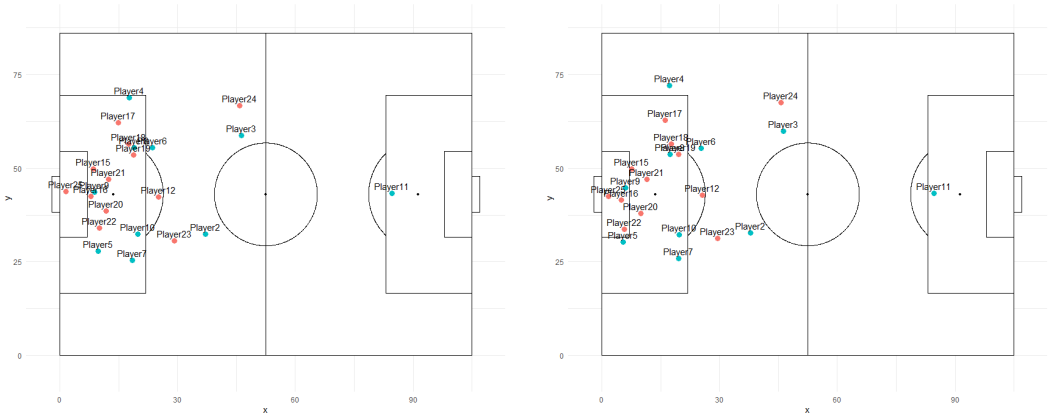
Também podemos observar que as ações do jogador 10 estão concentradas no canto inferior esquerdo do gráfico, o que indica que as ações que envolveram este atleta com a posse de bola concentram muitos jogadores ao seu redor.

type	**subtype**	κ	σ^2	ρ
PASS	CROSS	13.27544	0.2247530	0.9980548
PASS	DEEP BALL	14.17987	0.1549975	0.9984420
PASS	GOAL KICK	39.65578	0.7069947	0.9987440
PASS	HEAD	39.42690	0.3876083	0.9986822
PASS	THROUGH BALL-DEEP BALL	20.77655	0.3402594	0.9989782
PASS	NA	24.39029	0.4527520	0.9983016
SHOT	HEAD-ON TARGET-GOAL	12.79613	0.2192451	0.9979106
SHOT	HEAD-ON TARGET-SAVED	13.03576	0.1428571	0.9980398
SHOT	OFF TARGET	18.56709	0.9863079	0.9972241
SHOT	OFF TARGET-OUT	20.79802	0.3936425	0.9985049
SHOT	ON TARGET-GOAL	15.59383	0.2719920	0.9989326

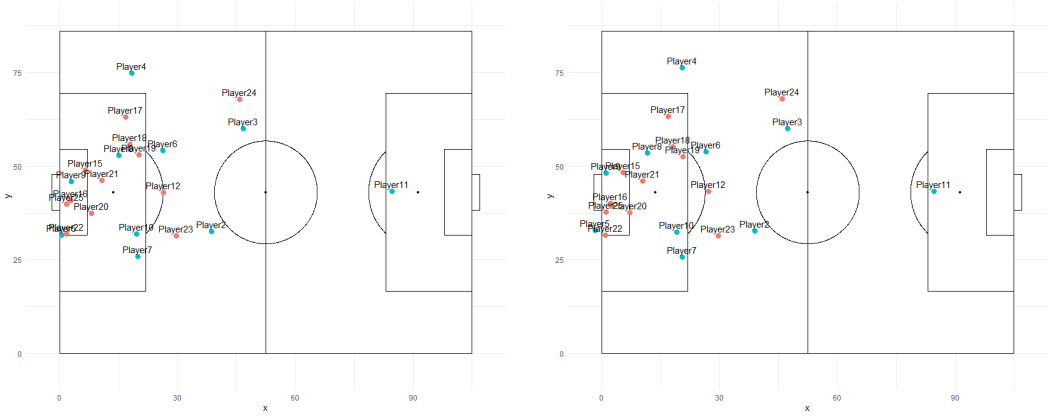
Tabela 4.2: Tabela resumo de tipos diferentes de ações

Na Figura 4.8 podemos observar no eixo x a range κ e no eixo y o desvio padrão da funçao de matern estimadas para cada jogador. Como se observa o jogador 2, um zagueiro, tem maior variação de *ranges* para suas jogadas, natural pelo contexto de jogos onde sabemos que zagueiros podem ficar trocando passes entre si sem importunação do adversário, bem como podem participar de situações onde tem que se desfazer da bola para evitar um perigo de gol do adversário.

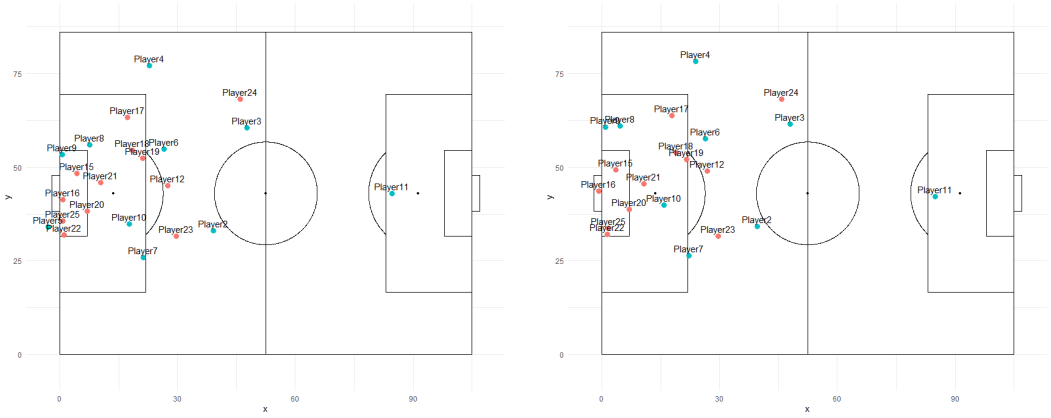
Na Figura 4.9 temos o mesmo gráfico que tinha como referência os jogadores e agora passamos a analisar o tipo de ação que gerou. O formato dos pontos é o tipo de ação e as cores são os subtipos correspondentes. Como mencionado anteriormente, podemos observar que situações de finalização (*Shot*) ocorrem em uma região com menor dependência espacial, o que indica que nessas situações há uma grande quantidade de atletas cujo posicionamento acabam sendo influenciadas por esta ação.



(a) *3959s* (b) *3960s*



(c) *3961s* (d) *3962s*



(e) *3963s* (f) *3964s*

Figura 4.7: Posição dos jogadores no momento do gol (*3962s*)

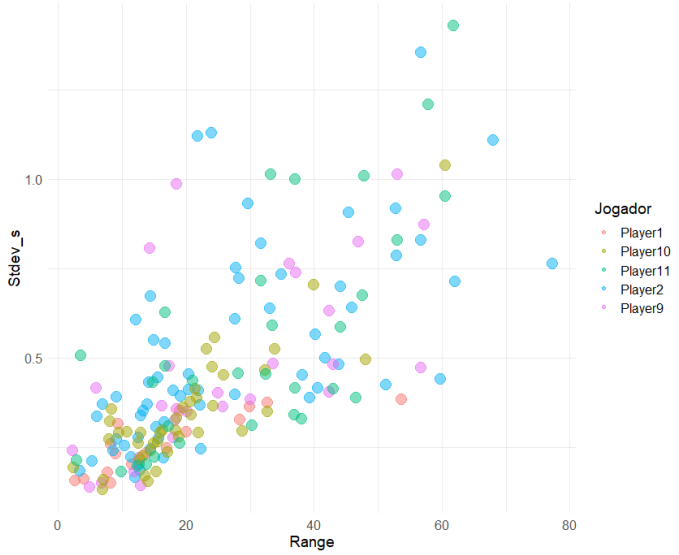


Figura 4.8: Gráfico de dispersão dos parâmetros σ e κ estimados para os diferentes jogadores.

Por fim, achamos interessante apresentar as curvas das funções de covariância de Matérn aplicadas a cada jogador. Nota-se que o goleiro (*Player11*) e zagueiro (*Player2*), as curvas de dependência espacial possuem uma range menor, indicando que a dependência espacial diminui mais rapidamente do que para as outras posições. Este resultado sugere uma menor correlação entre as ações desses jogadores e as de outros membros da equipe, o que pode indicar uma maior independência de suas atividades em relação ao restante do time. Isso pode ser atribuído às diferentes posições e funções desempenhadas por esses jogadores dentro do campo.

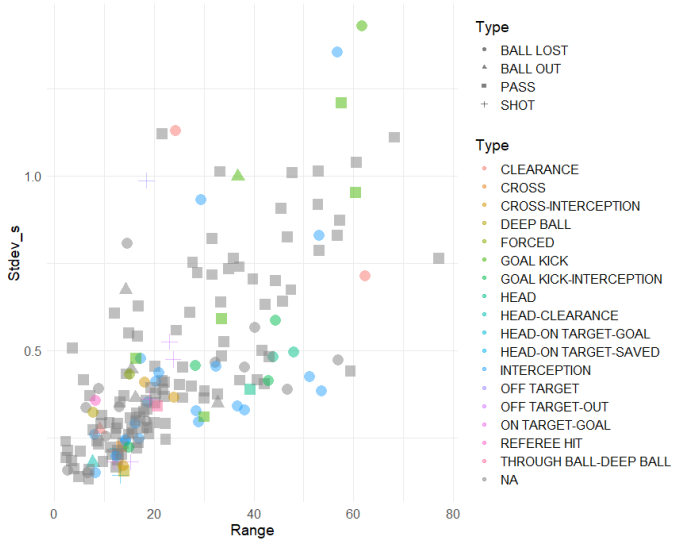


Figura 4.9: Gráfico de dispersão dos parâmetros σ e κ estimados para os diferentes jogadores dado o tipo de ação.

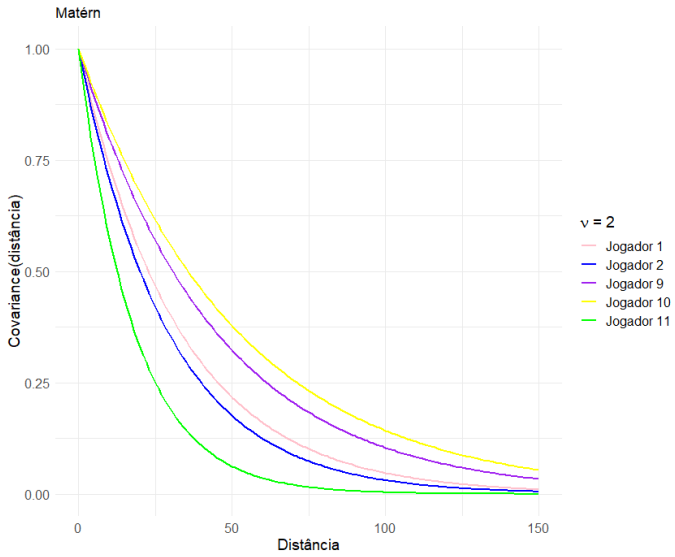


Figura 4.10: Gráfico funçao de covariância de Matérn estimadas para os diferentes jogadores.

5 Considerações finais

A ideia da presente pesquisa em explorar a aplicação de um modelo espaço temporal utilizando SPDE em dados de *tracking data* para a modelagem das zonas de influência de jogadores de futebol pode ser considerada inovadora dentro do esporte. Com os resultados, confirmamos algumas das suposições em relação a posições e influências que alguns atletas exercem em campo. Goleiros e zagueiros tem dependência espacial maior que os atacantes, o que é coerente com o que vemos em campo em partidas de futebol.

No entanto, é crucial reconhecer as limitações que encontramos ao desenvolver este trabalho. A principal delas reside na restrição ao acesso aos tipo de dado, um componente essencial para a precisão e abrangência das análises. O uso desse tipo de informação é de extrema relevância para compreender a dinâmica do jogo sem envolver necessariamente ações com a posse de bola, uma vez que detalhes cruciais sobre movimentação e interações entre jogadores podem ser perdidos.

Além disso, a censura das informações dos times, mesmo que seja uma prática comum em muitos contextos esportivos devido a proteção de dados e e direitos de imagem dos envolvidos, representa um desafio significativo. A falta de transparência em relação a certos dados impede uma interpretação objetiva e completa sobre o desempenho de jogadores específicos, limitando, assim, a extensão das conclusões que podem ser extraídas.

No momento que tivermos posse de dados que envolvam jogadores como Messi ou Suárez, podemos trazer análises com maior contexto de forma a acrescentar maior certeza nas conclusões que forem geradas.

Apesar destas limitações, acreditamos que o modelo oferece contribuições na capacidade de modelar as zonas de influência dos jogadores de forma espaço temporal. Essa abordagem não apenas permite uma avaliação mais abrangente das interações entre os jogadores, mas também pode informar estratégias táticas mais refinadas e servir como base comparativa de jogadores.

Referências Bibliográficas

Basketball-Reference (2023). Calculating per. `https://www.basketball-reference.com/about/per.html`. Acesso em 02/02/2024.

Blangiardo, M. e Camelatti, M. (2015). *Spatial and Spatio-temporal Models with R-INLA*. John Wiley & Sons, Ltd.

Confederação Brasileira de Futebol - CBF (2023). Em alta: Cbf registrou 1.276 clubes em 2022. `https://www.cbf.com.br/a-cbf/informes/index/em-alta-cbf-registrou-1-276-clubes-em-2022`. Acesso em 31/01/2024.

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Ltd.

Custódio, I. (2011). Análise quantitativa e qualitativa da carga de treinamento de uma equipe de futebol.

Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F., Bachl, N., e Pigozzi, F. (2006). Performance characteristics according to player position in elite soccer. *International Journal of Sports Medicine*.

Fernandez, J. e Bornn, L. (2018). Wide open spaces: A statistical technique for measuring space creation in professional soccer. *MIT Sloan Sports Analytics Conference*.

Fernández, J. (2021). *A framework for the analytical and visual interpretation of complex spatiotemporal dynamics in soccer.* PhD thesis, Polytechnic University of Catalonia.

Fédération Internationale de Football Association - FIFA (2023). One month on: 5 billion engaged with the fifa world cup qatar 2022™. `https://www.fifa.com/tournaments/mens/worldcup/qatar2022/news/one-month-on-5-billion-engaged-with-the-fifa-world-cup-qatar-2022-tm`. Acesso em 04/02/2024.

Gotway, C. e Waller, L. (2004). *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons, Ltd.

Kovalchik, S. (2022). Player tracking data in sports. *Annual Review of Statistics and Its Application*.

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., e Rue, H. (2019). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman & Hall/CRC.

Kullowatz, M. (2020). Goals added: Deep dive methodology. `https://www.americansocceranalysis.com/home/2020/5/4/goals-added-deep-dive-methodology`. Acesso em 26/01/2024.

Lee, P. (2012). *Bayesian Statistics: An Introduction, 4th Edition*. John Wiley & Sons, Ltd.

Liu, G., Luo, Y., Schulte, O., e Kharrat, T. (2020). Deep soccer analytics: learning an action-value function for evaluating soccer players. *Data Mining and Knowledge Discovery*.

Lunn, D., Thomas, A., Best, N., e Spielgelhalter, D. (2000). Winbugs: a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*.

Metrica-Sports (2024). `https://metrica-sports.com/automated-tracking-data/`. Acesso em 24/01/2024.

Moraga, P. (2019). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Chapman & Hall/CRC Biostatistics Series.

Nelder, J. e Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*.

Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. *The Open Sports Sciences Journal*.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. Version 4.3.2 - R Foundation for Statistical Computing, Vienna, Austria. Acesso em 26/02/2024.

R-INLA Project (2024). *R-INLA*. Acesso em 26/02/2024.

Reuters (2024). Premier league clubs spend record 2.36 billion pounds in transfer window. `https://www.reuters.com/sports/soccer/premier-league-clubs-spend-record-236-billion-pounds-transfer-window-2023-09-02/`. Acesso em 04/02/2024.

Rue, H., Lindgren, F., e Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society, Series B*.

Rue, H., Martino, S., e Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society, Series B*.

Shaw, L. e Glickman, M. (2020). Dynamic analysis of team strategy in professional football. *Barça Sports Analytics Summit*.

Sherman, M. (2011). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties (Wiley Series in Probability and Statistics Book 903)*. John Wiley & Sons, Ltd.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., e Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28.

Singh, K. (2018). Introducing expected threat (xt). `https://karun.in/blog/expected-threat.html`. Acesso em 28/01/2024.

Skillcorner (2024). `https://skillcorner.com/about`. Acesso em 24/01/2024.

Spearman, W. (2018). Beyond expected goals. *MIT Sloan Sports Analytics Conference*.

Stats-Perform (2019). Introducing a possession value framework. `https://www.statsperform.com/resource/introducing-a-possession-value-framework/`. Acesso em 26/01/2024.

Statsbomb (2021). Introducing on-ball value (obv). `https://statsbomb.com/articles/soccer/introducing-on-ball-value-obv/`. Acesso em 26/01/2024.

Sumpter, D. (2017). *Soccermatics: Mathematical Adventures in the Beautiful Game*. Bloomsbury Sigma.

Sykes, J. e Paine, N. (2016). How one man's bad math helped ruin decades of english soccer. `https://fivethirtyeight.com/features/how-one-mans-bad-math-helped-ruin-decades-of-english-soccer`. Acesso em 16/01/2024.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*.

Tsokos, A., Narayanan, S., Kosmidis, I., Baio, G., Cucuringu, M., Whitaker, G., e Király, F. (2018). Modeling outcomes of soccer matches. *Machine Learning (2019) 108:77–95*.

Wired (2020). The nba's game-changing approach to data. `https://www.wired.com/sponsored/story/the-nbas-game-changing-approach-to-data/`. Acesso em 02/02/2024.

Anexos

Nas tabelas abaixo, temos as estimativas das modas das distribuições *a posteriori* calculadas para cada um dos parâmetros da função de covariância de matern, dado as diferentes ações envolvendo os atletas mencionados.

- *event* - é a indexação do evento referente apenas aos eventos que o jogador participou.
- (κ) *range* - é o parâmetro de que indica a partir de qual distância a ação efetuada pelo atleta não possui mais dependência espacial.
- σ *stdev s* - é o parâmetro que mede a variabilidade referente à distância
- ρ *group rho* - é referente à correlação temporal. Quanto mais próximo de 1 maior a correlação.
- *type* - tipo de evento
- *subtype* - subtipo de evento

event	**range (κ)**	**stdev_s (σ)**	**group_rho (ρ)**	**type**	**subtype**
1	32.750253	0.3739121	0.9993862	PASS	NA
2	18.845575	0.3517901	0.9994788	BALL LOST	INTERCEPTION
3	16.936897	0.2471234	0.9978165	BALL LOST	INTERCEPTION
4	53.833717	0.3835899	0.9990998	BALL LOST	INTERCEPTION
5	13.784079	0.2337060	0.9981840	PASS	NA
6	7.729668	0.1789095	0.9977894	BALL OUT	HEAD-CLEARANCE
7	7.993739	0.2595800	0.9990563	BALL LOST	INTERCEPTION
8	9.060306	0.2312464	0.9984748	PASS	NA
9	20.017218	0.2931984	0.9989622	PASS	NA
10	29.869033	0.3638810	0.9989114	PASS	NA
11	28.534310	0.3272955	0.9986464	BALL LOST	INTERCEPTION
12	8.129400	0.1498798	0.9969880	BALL LOST	INTERCEPTION
13	3.947805	0.1608421	0.9980927	PASS	NA
14	9.396276	0.3145907	0.9986660	PASS	NA
15	2.584126	0.1572191	0.9949587	BALL LOST	NA
16	11.444399	0.2026063	0.9961743	PASS	NA

Tabela 5.1: Saída do modelo com jogadas envolvendo jogador 1

event	**range (κ)**	**stdev_s (σ)**	**group_rho (ρ)**	**type**	**subtype**
1	51.333414	0.4237333	0.9994969	BALL LOST	INTERCEPTION
2	39.426904	0.3876083	0.9986822	PASS	HEAD
3	59.564389	0.4397945	0.9992227	PASS	NA

Tabela 5.2: Continuação da Tabela 5.2

event	**range** (κ)	**stdev_s** (σ)	**group_rho** (ρ)	**type**	**subtype**
4	12.772495	0.2030214	0.9976855	PASS	NA
5	22.441625	0.2454382	0.9989081	PASS	NA
6	3.506712	0.1849094	0.9978723	PASS	NA
7	43.856960	0.4818146	0.9994832	BALL LOST	HEAD
8	40.045992	0.5656487	0.9999001	BALL LOST	NA
9	9.168457	0.2727717	0.9981126	BALL OUT	CLEARANCE
10	45.855793	0.6405687	0.9993319	PASS	NA
11	13.137617	0.3513819	0.9971957	PASS	NA
12	14.281487	0.2450985	0.9982114	BALL LOST	INTERCEPTION
13	11.332117	0.2218230	0.9954783	PASS	NA
14	41.715573	0.4992586	0.9996131	PASS	NA
15	22.105290	0.4091929	0.9977505	PASS	NA
16	16.506759	0.3205401	0.9983263	PASS	NA
17	8.732162	0.2417192	0.9992845	PASS	NA
18	12.222444	0.1660156	0.9980698	PASS	NA
19	15.189291	0.3070622	0.9990280	PASS	NA
20	27.660726	0.3983697	0.9985784	PASS	NA
21	15.721453	0.2744187	0.9983591	PASS	NA
22	6.144029	0.3360695	0.9982102	BALL LOST	NA
23	10.404753	0.2551913	0.9991843	BALL LOST	NA
24	14.303277	0.6722287	0.9981044	BALL OUT	NA
25	12.521658	0.2773855	0.9979958	PASS	NA
26	33.268763	0.6390652	0.9997208	PASS	NA
27	16.521834	0.2198675	0.9976874	PASS	NA
28	20.353257	0.4100455	0.9991091	BALL LOST	INTERCEPTION
29	20.421135	0.4542551	0.9999432	PASS	NA
30	13.127385	0.3390930	0.9988732	BALL LOST	NA
31	17.989086	0.4098434	0.9987024	BALL LOST	CROSS-INTERCEPTION
32	38.271942	0.4514626	0.9995856	BALL LOST	NA
33	15.653124	0.4449926	0.9993020	BALL OUT	NA
34	11.980473	0.6067499	0.9988183	PASS	NA
35	8.998724	0.3897277	0.9992014	BALL LOST	NA
36	22.120058	0.3681682	0.9987316	PASS	NA
37	19.234072	0.3921625	0.9956557	PASS	NA
38	7.069784	0.3701030	0.9989925	PASS	NA
39	13.832908	0.3693896	0.9974121	PASS	NA
40	40.504824	0.4161167	0.9994647	PASS	NA
41	5.401380	0.2110291	0.9975401	PASS	NA
42	12.640217	0.1861253	0.9987404	PASS	NA
43	62.295342	0.7132345	0.9989754	BALL LOST	CLEARANCE
44	29.572318	0.9312876	0.9992695	BALL LOST	INTERCEPTION
45	15.013512	0.5497420	0.9998837	PASS	NA
46	16.547680	0.5395873	0.9992612	PASS	NA
47	44.217315	0.6986784	0.9980493	PASS	NA
48	77.224856	0.7635852	0.9998306	PASS	NA
49	27.412961	0.6094389	0.9997608	PASS	NA
50	34.866786	0.7339711	0.9976625	PASS	NA
51	52.915103	0.7853356	0.9979010	PASS	NA
52	21.716924	1.1189748	0.9956476	PASS	NA
53	14.341778	0.4325582	0.9978821	PASS	NA
54	56.959198	1.3531888	0.9999042	BALL LOST	INTERCEPTION
55	56.800554	0.8296870	0.9976717	PASS	NA
56	27.808503	0.7517405	0.9995229	PASS	NA
57	68.105293	1.1092030	0.9974427	PASS	NA
58	52.858212	0.9182543	0.9990126	PASS	NA
59	45.431724	0.9051675	0.9975693	PASS	NA
60	28.506885	0.7226281	0.9974408	PASS	NA
61	31.833287	0.8197974	0.9959813	PASS	NA
62	24.179911	1.1284355	0.9966739	BALL LOST	CLEARANCE

Tabela 5.2: Saída do modelo com jogadas envolvendo jogador 2

event	range (κ)	stdev_s (σ)	group_rho (ρ)	type	subtype
1	18.008854	0.2770389	0.9985322	PASS	NA
2	12.796133	0.2192451	0.9979106	SHOT	HEAD-ON TARGET-GOAL
3	25.110351	0.4011958	0.9991439	PASS	NA
4	5.078549	0.1388630	0.9974764	PASS	NA
5	42.348757	0.4042829	0.9990443	PASS	NA
6	6.716976	0.1499451	0.9970732	BALL LOST	NA
7	20.190827	0.3504821	0.9975545	PASS	NA
8	42.234926	0.6315408	0.9989872	PASS	NA
9	43.187191	0.4818901	0.9985630	PASS	NA
10	30.029100	0.3837040	0.9989304	PASS	NA
11	25.674866	0.3644890	0.9992476	PASS	NA
12	56.758518	0.4729416	0.9990531	BALL LOST	NA
13	16.283331	0.3647602	0.9987207	BALL OUT	NA
14	18.128173	0.3240441	0.9985847	PASS	NA
15	14.468831	0.8055435	0.9993284	BALL LOST	NA
16	13.035758	0.1428571	0.9980398	SHOT	HEAD-ON TARGET-SAVED
17	57.308840	0.8718320	0.9977599	PASS	NA
18	18.396250	0.3556412	0.9986014	PASS	NA
19	11.883826	0.1827998	0.9990464	SHOT	ON TARGET-GOAL
20	2.145564	0.2397004	0.9982578	PASS	NA
21	5.875830	0.4157486	0.9932610	PASS	NA
22	33.587869	0.4837263	0.9982082	PASS	NA
23	36.018596	0.7637779	0.9988207	PASS	NA
24	53.072460	1.0131966	0.9985702	PASS	NA
25	17.428496	0.4778356	0.9987862	BALL LOST	INTERCEPTION
26	46.847984	0.8246519	0.9965197	PASS	NA
27	37.225473	0.7391010	0.9987147	PASS	NA
28	18.567092	0.9863079	0.9972241	SHOT	OFF TARGET

Tabela 5.3: Saída do modelo com jogadas envolvendo jogador 9

event	range (κ)	stdev_s (σ)	group_rho (ρ)	type	subtype
1	28.773927	0.2957666	0.9988652	BALL LOST	INTERCEPTION
2	7.170262	0.1587296	0.9955691	PASS	NA
3	13.275444	0.2247530	0.9980548	PASS	CROSS
4	18.502629	0.2976358	0.9984916	BALL LOST	NA
5	15.312965	0.1809574	0.9980476	SHOT	OFF TARGET-OUT
6	47.988934	0.4946078	0.9996029	BALL LOST	HEAD
7	14.179866	0.1549975	0.9984420	PASS	DEEP BALL
8	16.995538	0.2361960	0.9983085	PASS	NA
9	6.930191	0.1308507	0.9976595	PASS	NA
10	18.516918	0.3320152	0.9983779	PASS	NA
11	20.481141	0.3773353	0.9993956	PASS	NA
12	32.259989	0.4662337	0.9996163	BALL LOST	NA
13	32.840073	0.3493208	0.9978910	BALL OUT	NA
14	33.915569	0.5255867	0.9987190	PASS	NA
15	21.439454	0.4134650	0.9993878	PASS	NA
16	2.293361	0.1932778	0.9992542	PASS	NA
17	20.776546	0.3402594	0.9989782	PASS	THROUGH BALL-DEEP BALL
18	21.663525	0.3882257	0.9991425	PASS	NA
19	25.888932	0.4511285	0.9994000	PASS	NA
20	12.530278	0.2606989	0.9992787	PASS	NA
21	23.968362	0.4744610	0.9992307	SHOT	OFF TARGET-OUT
22	16.107755	0.2918177	0.9972874	BALL LOST	INTERCEPTION
23	14.359731	0.2411968	0.9979522	BALL LOST	INTERCEPTION
24	16.152408	0.2984718	0.9984944	PASS	NA
25	15.107104	0.2601656	0.9979722	PASS	NA
26	18.778486	0.2801726	0.9969766	PASS	NA
27	24.387378	0.5574923	0.9989375	PASS	NA
28	12.820492	0.2906166	0.9988102	PASS	NA
29	13.749983	0.1697888	0.9988194	BALL LOST	CROSS-INTERCEPTION
30	7.826865	0.2725703	0.9961930	PASS	NA
31	60.665000	1.0382908	0.9975248	PASS	NA
32	15.703713	0.2646286	0.9984261	PASS	NA
33	19.303834	0.3611843	0.9988187	SHOT	ON TARGET-GOAL
34	7.972244	0.3217201	0.9992985	BALL LOST	DEEP BALL
35	24.096199	0.3646905	0.9980302	BALL LOST	CROSS-INTERCEPTION
36	10.922122	0.2929410	0.9975779	PASS	NA
37	40.027311	0.7039121	0.9993123	PASS	NA
38	12.742146	0.2147157	0.9980152	PASS	NA
39	22.099047	0.2907520	0.9986402	PASS	NA
40	9.580759	0.2906323	0.9973027	PASS	NA
41	8.228008	0.3566648	0.9987618	BALL LOST	REFEREE HIT
42	23.112742	0.5255091	0.9982365	SHOT	OFF TARGET-OUT

Tabela 5.4: Saída do modelo com jogadas envolvendo jogador 10

event	range (κ)	stdev_s (σ)	group_rho (ρ)	type	subtype
1	44.198074	0.5850603	0.9991375	BALL LOST	GOAL KICK-INTERCEPTION
2	30.230279	0.3101318	0.9990187	PASS	GOAL KICK
3	12.338897	0.1982070	0.9977215	BALL LOST	INTERCEPTION
4	37.099659	0.4145163	0.9987769	PASS	NA
5	42.997072	0.4143460	0.9990438	BALL LOST	GOAL KICK-INTERCEPTION
6	36.868044	0.3397012	0.9994652	BALL LOST	INTERCEPTION
7	14.813884	0.2215934	0.9990434	BALL LOST	GOAL KICK-INTERCEPTION
8	17.312394	0.3094925	0.9989845	PASS	NA
9	38.099529	0.3301103	0.9995102	BALL LOST	INTERCEPTION
10	46.700429	0.3891261	0.9992084	BALL LOST	NA
11	19.108011	0.2615343	0.9968892	PASS	NA
12	16.516074	0.4770589	0.9989931	PASS	GOAL KICK
13	2.859713	0.2135192	0.9995291	PASS	NA
14	9.918771	0.1824049	0.9983102	PASS	NA
15	3.487657	0.5066528	0.9997511	PASS	NA
16	15.070758	0.4326091	0.9997977	BALL LOST	FORCED
17	21.037670	0.4371034	0.9997250	BALL LOST	INTERCEPTION
18	28.303463	0.4573973	0.9992125	BALL LOST	GOAL KICK-INTERCEPTION
19	47.544852	0.6735047	0.9995122	PASS	NA
20	16.892035	0.6270381	0.9980567	PASS	NA
21	13.612075	0.2022444	0.9976823	PASS	NA
22	32.372015	0.4531936	0.9993676	BALL LOST	INTERCEPTION
23	61.827393	1.4275978	0.9989092	BALL LOST	GOAL KICK
24	53.224127	0.8285557	0.9979960	BALL LOST	INTERCEPTION
25	33.149314	1.0117346	0.9978123	PASS	NA
26	57.675416	1.2073695	0.9999971	PASS	GOAL KICK
27	31.735435	0.7163972	0.9990930	PASS	NA
28	60.325846	0.9506981	0.9985151	PASS	GOAL KICK
29	47.690451	1.0086293	0.9993406	PASS	NA
30	33.531273	0.5897152	0.9971960	PASS	GOAL KICK
31	36.979267	0.9983701	0.9998399	BALL OUT	GOAL KICK

Tabela 5.5: Saída do modelo com jogadas envolvendo jogador 11